THE EARTH

Cambridge University Press
Fetter Lane, London

New York
Bombay, Calcutta, Madras
Toronto
Macmillan

Tokyo
Maruzen-Kabushiki-Kaisha

THE EARTH

ITS ORIGIN, HISTORY AND PHYSICAL CONSTITUTION

By
HAROLD JEFFREYS, M.A., D.Sc., F.R.S.

SECOND EDITION

CAMBRIDGE : AT THE
UNIVERSITY PRESS
MCMXXIX

PRINTED IN GREAT BRITAIN

TO

THE MEMORY OF THE LATE

SIR GEORGE HOWARD DARWIN

THE FOUNDER OF MODERN
COSMOGONY AND GEOPHYSICS

PREFACE

During the years 1920–23 I delivered three times in this college a course of eight lectures on the Physics of the Earth's Interior. The course was intended to give an outline of present knowledge concerning what may be called the major problems of geology, namely the physical constitution of the earth, the causes of mountain formation, and the nature of isostasy. It was, however, impossible to give in the lectures full accounts of the arguments employed, partly because the course was too short, and partly because the mathematical knowledge of the listeners was extremely varied. Accordingly this book has been written; the argument has been made as connected as appeared possible, and various geophysical topics that could not be discussed in the lectures, such as the variation of latitude, have been introduced. The aim has been to discuss the theories of the main problems of geophysics, and to exhibit their interrelations. Several large branches have, however, been almost entirely omitted; terrestrial magnetism, atmospheric electricity, tidal theory (apart from tidal friction) and meteorology have received little or no attention, because to give anything approaching an adequate account of any of them would have required a longer discussion than their connexion with the original topics of the book seemed to warrant.

I have attempted to describe the present position of the subject, rather than its history. For this reason several pieces of work of capital importance in the development of geophysics have escaped mention. Lord Kelvin's estimate of the rigidity of the earth from its tidal yielding is a case in point; it has not been discussed because more detailed and definite information can now be derived from seismology. Sir G. H. Darwin's pioneer work on tidal friction, again, has been only outlined, because it was mainly a discussion of bodily tides in a homogeneous earth, which now appear to be comparatively unimportant in influencing the evolution of the earth and moon. Nevertheless if the contents of the second volume of Darwin's collected papers had not been published, it is improbable that Chapters III and XIV of this book would have been written.

Quantitative comparison of theory with fact has always been the main object of the book, and practically all the theories advocated have survived the test of quantitative application to several phenomena. Accurate theories have been given where they seemed necessary for this purpose; but where an estimate of an order of magnitude was all that was required, and could be obtained by rough methods, such methods have always been used. I have been encouraged in the latter course by several facts. First, though the method of orders of magnitude is not convincing to the pure

mathematician, it is a matter of experience that when a problem discussed by this method is afterwards solved by more formal methods, the answer is found to be of the correct order of magnitude, which is all that the method claims; it could be vitiated only by a fortuitous numerical coincidence. Second, though in some cases formally accurate solutions of related problems exist, or could be obtained, the problems actually so soluble differ so much from those that actually arise in geophysics that, in their actual application, they could at best be correct only as regards order of magnitude. Thus they are not inherently any more informative than the rougher methods. Third, a direct proof that a particular hypothesis will account for particular data is not very strong confirmation of the hypothesis when both the data and the consequences of the hypothesis are known only vaguely; but if it is shown that the results of the hypothesis agree with the facts as regards order of magnitude, while the results of denying it are in definite disagreement, the confirmation of the hypothesis will be almost as strong as if a close agreement had been obtained. The method of exhaustion of alternatives is specially useful in geophysics, because incorrect geophysical hypotheses usually fail by extremely large margins.

Two criticisms are certain to be made by geologists, and therefore I venture to attempt to meet them in advance. The first is that the book contains a great deal of matter not of a geophysical character. I thought at first of avoiding this objection by dividing the book into two parts, one cosmogonical and one geophysical; but I found such a course impossible, since the two were too closely interwoven, each depending in part on the results of the other. In a work mainly theoretical rather than descriptive in character it therefore seemed best to develop the implications of a hypothesis wherever they might lead to results capable of empirical test, rather than to confine my attention to one particular planet. If a theory is satisfactory, the more it is shown to explain the more reliable it is; and if it is unsound, it will be unsound whether the fact inconsistent with it happens to relate to the earth or to the satellites of Uranus.

The other objection I anticipate is that the book is too mathematical for geological readers. The answer is simple: the results aimed at are quantitative, and there is no way of obtaining quantitative results without mathematics. I have tried to keep the mathematics as elementary as possible; but some problems could not be handled by simple mathematics, and I had no alternative except to give all that was necessary. If the geologist cannot follow a part of the book, I hope he will omit it and go on to the next non-mathematical passage, trusting that someone else will point out any intervening mistakes (and the mathematically trained readers, with few exceptions, will do just the same). He will then know that at any rate some people will be able to follow the argument all through, and he will see just where he fails to follow it himself; whereas

a so-called non-mathematical "exposition" would only bewilder the mathematical physicist, while making it impossible for the geologist to see how much is hypothesis and how much is merely the systematic investigation of the consequences of hypotheses already made and data already found. In short, if geophysics requires mathematics for its treatment, it is the earth that is responsible, not the geophysicist.

The paragraphs are numbered according to the decimal system; of any two paragraphs, that with the smaller number comes earlier in the book. The integral part of a paragraph number is the number of the chapter. Equations have as a rule been numbered consecutively through each paragraph; but in some cases they have been numbered consecutively through several closely related paragraphs. In cross-references, where reference is made to another equation in the same paragraph, only the number of the equation is given; but where the reference is to a different paragraph, the numbers of that paragraph and of the equation are given; *e.g.* 14·61 (3).

I wish to express my thanks to the staff of the Cambridge University Press for their courtesy during the publication of this book; to Mr R. Stoneley, who has read the whole in proof and checked a great deal of the mathematical work, suggesting many improvements in the process; to Dr J. H. Jeans, who verified Chapter IX; to Dr A. A. Griffith, who gave me much of the information incorporated in Chapter XI, though he does not wholly approve of my terminology; to Dr Arthur Holmes, whose influence on my geophysical thought has been none the less important because I experienced it before beginning this book; and to Sir Ernest Rutherford, Prof. Eddington, Prof. Shapley, and Dr Wrinch, who have read various portions in typescript and suggested improvements.

HAROLD JEFFREYS

ST JOHN'S COLLEGE
CAMBRIDGE

1924 *March* 6

PREFACE TO THE SECOND EDITION

A NEW edition of this work must begin with an expression of thanks to numerous reviewers for their kind notices of the first, and to the correspondents that have suggested emendations of it. The latter are, I fear, too many to be thanked individually.

The revision has required the rewriting and rearrangement of a large part of the book. Seismology, which occupied a somewhat subordinate place, has had to be much expanded, mainly as a result of important work done on the Continent between 1907 and 1916, which only came to my notice after the publication of the first edition. The other alterations have arisen in an attempt to keep up with the growth of the subject during the last four years. Eddington's revised theory of stellar constitution has led to some quantitative changes in the account of the early history of the planets; recent developments in seismology, in the study of both near earthquakes, surface waves, and the bodily waves from distant earthquakes, have made it possible to go into much more detail about the structure of the earth. It has also become possible to be much more definite than before about the nature and degree of isostatic compensation. More attention has been given to the bodily tide, which proved to supply an essential datum concerning the mechanics of the interior, a datum that seismology had shown to be required, but had not itself supplied. The account of the earth's thermal history has been recast. The portions with geological applications have been placed together in Chapter XV, and considerably expanded.

The gist of the papers awarded the Adams Prize of the University of Cambridge in 1927 has been included. As the papers have all appeared in well-known journals complete republication was thought unnecessary. The present revision was done in April to June 1928, but it has in some cases been possible to refer to later work in passages added during passage through the press.

Thanks are again due to the staff of the Cambridge University Press for their care and consideration during publication, and to Mr R. Stoneley and Prof. A. Holmes for valuable criticisms and assistance in proof-correction. Acknowledgments for information and other help have as far as possible been indicated in the text. Prof. R. A. Daly has also placed me under special indebtedness by directing my attention to new aspects of problems under discussion. Dr C. E. Tilley has kindly checked the information at the beginning of Chapter VI.

HAROLD JEFFREYS

ST JOHN'S COLLEGE
CAMBRIDGE
1928

INTRODUCTION

"I am doing the best I can at my age."
G. B. SHAW, Preface to *Back to Methuselah*.

THE special difficulty of the problem of the physics of the earth's interior arises from the extremely restricted range, both of depth and of time, for which any direct evidence is available. The deepest boring yet made is one of 2·5 km., and there are few over a kilometre in depth; mines are still shallower; and apart from borings and mines the only accessible parts of the crust are those exposed in quarries and cliffs, and the land surface itself. Samples of material from greater depths are continually being brought up to the surface, but the possibility of utilizing any information they may give us is seriously hampered by the fact that we know only vaguely from what depths they have come and to what extent their temperature and even their chemical constitution have been altered in their ascent. In addition, all observations of earthquakes and all observations of topography and gravity have to be made at the surface or near it. Thus the problem of the physics of the earth's interior is to make physical inferences over a range of depth of over 6000 km. from data determined only for a range of 2 km. at the outside.

Extrapolation over such a range can be carried out only if we have some extremely reliable scientific laws, well verified in the laboratory and the observatory, to guide us in our investigation. Fortunately this is the case. Considerable headway can be made by means of such laws as those of gravitation, heat conduction, radioactive disintegration, elasticity, and fluid friction, all of which are well established. It is naturally impossible to proceed far in such a subject without introducing, besides known laws and empirical data, some *ad hoc* assumptions. The number of these has to be restricted as severely as possible; the justification of any hypothesis lies in its ability to coordinate otherwise unrelated facts, and hypotheses that do not explain more than they assume should be omitted.

In many questions relating to the interior of the earth some knowledge of its thermal state is indispensable. There is no direct method of arriving at this knowledge; but we may make use of the laws of heat conduction to arrive at it indirectly. Given the distribution of temperature in a solid at any instant, and the thermal conductivity, it is possible to find the distribution at any subsequent instant, and therefore if we know the thermal conditions inside the earth at some past date it should be possible, given sufficient mathematical labour, to find those of the present time. We have not such knowledge for any historical date, but we can find it for the time when the earth was first formed, by making use of the modern

theories of the origin of the solar system. It is therefore necessary to give first of all an account of the evidence in favour of these theories; but in so doing we must make use of information relating to the other members of the solar system besides the earth. It is found that the theory finally selected as the most satisfactory gives fairly definite information about the initial thermal state of the earth, which we can use in determining its subsequent history.

The thermal history of the earth is influenced not only by the initial distribution of temperature, but also by radioactive substances in the crust, which are responsible at the present time for the greater part of the heat that is continually being conducted up to the surface and finally radiated away. It is therefore necessary to make a great deal of use of the information supplied by geologists concerning radioactivity in rocks, especially in relation to the age of the older rocks and the present thermal output. The thermal history in its turn supplies information about the elastic properties of the interior of the earth and about the deformations set up by contraction and other disturbing factors. In this way it is found possible to give a coherent account of the dynamics of isostasy, in relation to the support of both continents and mountains.

We wish, then, to arrive at a decision concerning the primitive state of the earth in order to utilize it in an account of the subsequent history and of the present physical condition. The chief questions to be dealt with are first, whether the earth was ever mostly fluid, or whether it has always been solid throughout, and cool near the surface; and second, if it be granted that it was ever fluid, to form some conception of the manner of solidification and the physical conditions just after solidification. A direct answer to either of these questions requires a discussion of the origin of the earth, concerning which numerous conflicting hypotheses have been propounded. A stage in the development of cosmogony has, however, now been reached in which both points can be decided with a high degree of probability.

Practically all hypotheses regarding the origin of the solar system agree in supposing the sun and the planets to have formerly formed a single mass; they differ in the processes supposed to have led to the separation of the mass into many parts and in the subsequent evolution of the planets. Most of the hypotheses may be divided into two main groups, which may be called the Rotational and Tidal types. In the rotational type the rupture is supposed to have arisen through the speed of rotation increasing, as the mass condensed, to such an extent that gravity ceased to be adequate to hold it together. In the discussion of this type the leading investigators have been Laplace, Babinet, Roche, Moulton and Jeans. In the tidal type a star larger than the sun is supposed to have come close to the sun and practically torn it asunder by its tidal action. A hypothesis resembling this was advanced by Buffon, but fell out of

favour after the criticisms of Laplace. Its modern development was due in the first place to Chamberlin and Moulton; but several features of their formulation have been found unsatisfactory by the present writer, who has therefore attempted to reconstruct the theory entirely from the time of the rupture. At the same time Jeans has carried out a comprehensive and general dynamical discussion of the possible methods of break-up of fluid masses, and has developed in it an account of the processes concerned in the formation of a system by tidal action. The theory he has reached bears a close resemblance to mine, the agreement being especially striking since the two methods of attack are quite different. The theory adopted in this work will be of the tidal type, and based on a comparison of those already advanced by Jeans and myself. It is not, however, possible to give an adequate statement of the reasons for adopting this particular theory without first giving some account of the advantages and drawbacks of the earlier theories of the origin of the solar system. Descriptions will therefore be given of the two chief theories that preceded the present one, namely the Nebular Hypothesis of Laplace and Roche, and the Planetesimal Hypothesis of Chamberlin and Moulton. It has, however, been found more convenient to treat the latter in an Appendix than to place it in its proper historical position, which would have been immediately after the Laplacian theory.

The tidal theory of the origin of the solar system leads directly to the inference that the earth was formerly fluid, and hence to an account of its method of solidification and of the distribution of temperature immediately afterwards. The age of the earth and the amount of radioactive matter near the surface are found from modern information about radioactivity in rocks. In former works on the earth's thermal history since solidification it was good enough to proceed by means of simplifying hypotheses about the vertical distribution of radioactivity; but the information provided by modern seismology enables us now to go into greater detail. Seismology, with the information it gives concerning the mechanical properties of the earth all the way to the centre, is therefore discussed immediately after the age of the earth. The results are then combined for the discussion of the thermal history and incidentally the present distribution of temperature. It is shown that the observed phenomena of mountain building and isostasy are substantially what would be expected from this work. The evidence supplied by the figures of the earth and moon, tidal friction, the bodily tide, and the variation of latitude is discussed in other chapters.

Some topics of geophysical interest, though outside the main argument of the book, are discussed in appendices. These are the Planetesimal Theory, Jeans's Theory (points not discussed in Chapter II), the Relation of Geophysics to Geology, Theories of Glaciation, and Empirical Periodicities.

The main purpose of the book is to show how far the various modes of attack on the problems of geophysics have led to harmonious results. Accordingly information from many widely separated fields has had to be collected. Probably the range of the subjects discussed is so wide that nobody could possess a specialist's knowledge of all; I am quite certain that neither I nor anybody else possesses such knowledge at present. Imperfections are therefore unavoidable. Nevertheless I think that geophysics has suffered in the past from the lack of any systematic attempt to coordinate its various branches, and that many wild theories might have remained unpublished or, if published, have received only their fair share of attention, had it been possible to see at once how far they conflicted with the data of related branches of the subject. If therefore this book fails, as it must, to instruct the specialist in the data of his own subject, but sometimes enables him to check a hypothesis by means not only of his own data, but by some of those of related subjects, it will have served its purpose. Geophysics is no longer a field for uncontrolled speculation; it is a single science whose data are harmoniously coordinated by a definite physical theory, and any theory that contradicts this one must be shown to coordinate all these data equally satisfactorily before it can be accepted.

CHAPTER I

The Nebular Hypothesis of Laplace

"Though the mills of God grind slowly, yet they
grind exceeding small." LONGFELLOW.

1·1. *Laplace's Account of the Nebular Hypothesis.* The Nebular Hypothesis is usually attributed to Laplace, but the account of it given by him was exceedingly vague, no quantitative discussion whatever being included. It occurred in a semi-popular work, entitled *Exposition du système du monde,* the second edition of which was published in the seventh year of the French Revolution. The following account is a translation of that given by Laplace in the sixth chapter of the fifth book of this work:

Let us now pause to consider the arrangement of the solar system and its relations to the stars. The immense globe of the sun, the focus of the planetary movements, turns on its axis in twenty-five and a half days. Its surface is covered by an ocean of luminous matter whose active effervescences form variable spots, often very numerous, and sometimes larger than the earth. Above this ocean, a vast atmosphere rises; it is beyond it that the planets, with their satellites, move in almost circular orbits, in planes slightly inclined to the solar equator. Countless comets, after approaching the sun, depart to distances that prove that its empire extends much further than the known limits of the planetary system. Not only does the sun attract all these globes, leading them to move around itself, but also it sheds on them its light and its heat. Its beneficent action produces the growth of animals and plants on the earth, and analogy leads us to think that it produces similar effects on the planets; for it is not natural to think that the matter whose fertility we see developed in so many ways is sterile on so great a planet as Jupiter, which, like the terrestrial globe, has its days, its nights, and its years, and on which changes implying very vigorous activity are indicated by observation. Man, made for the temperature he enjoys on the earth, could not, to all appearance, live on the other planets; but must there not be endless organisms adapted to the various temperatures of the globes of this system? If the difference of elements and climates alone can make such variety in terrestrial productions, how much more must those of the various planets and their satellites differ? The most lively imagination can form no idea, but their existence is very probable.

Although the elements of the system of the planets are arbitrary, there are nevertheless very remarkable relations between them that may enlighten us with regard to their origin. It is astonishing to see that all the planets move around the sun from west to east, and almost in the same plane; all the satellites move around their primaries in the same direction and almost in the same plane as the planets; and finally, the sun, the planets and the satellites whose rotations have been observed turn on their axes in the same direction, and nearly in the plane of their revolutions in their orbits.

So extraordinary a phenomenon is not the result of chance; it indicates that a general cause has determined all these movements. To find approximately the probability with which this cause is indicated, we observe that the planetary

system as we know it to-day is composed of seven planets and eighteen satellites; the rotations of the sun, five planets, the moon, the satellites of Jupiter, Saturn's ring, and its most remote satellite, have also been observed. These movements together form a set of thirty-eight movements in the same direction, at least when one refers them to the plane of the solar equator, to which it seems natural to compare them. If we consider the plane of any direct movement, lying at first upon the plane of the sun's equator, and then becoming inclined to this last plane, and traversing all the degrees of inclination from zero to two right angles, it is clear that the movement will be direct for all inclinations less than a right angle and retrograde for greater inclinations. Thus all motions, direct and retrograde, can be represented by a change in inclination alone. The solar system, looked at from this point of view, offers then thirty-seven movements whose planes are inclined to that of the solar equator by at most a right angle. Supposing that their inclinations are due to chance, they could have extended to two right angles; and the probability that at least one of them would have exceeded one right angle would be $1 - \frac{1}{2^{37}}$ or $\frac{137438953471}{137438953472}$. It is then extremely probable that the direction of the planetary movements is not at all the effect of chance, and this becomes still more probable when we realize that the inclinations of the majority of these movements to the sun's equator are very small and much less than a right angle.

Another equally remarkable characteristic of the solar system is the smallness of the eccentricities of the orbits of the planets and their satellites, whereas those of the comets are much elongated. No intermediate stages between large and small eccentricities occur in the system. We are then again compelled to recognize the effect of a regular cause. Chance alone would not have given an almost circular form to the orbits of all the planets, and the cause that determined the movements of these bodies must therefore have rendered them nearly circular. This cause again must have influenced the great eccentricity of the orbits of the comets, and, what is very extraordinary, without having affected the directions of their movements; for when we consider the orbits of retrograde comets as inclined at more than a right angle to the plane of the ecliptic, we find that the mean inclination of the orbits of all observed comets is almost a right angle, as it would be if these bodies had been projected at random.

Thus, returning to the cause of the primitive movements of the planetary system, we have the five following facts:

1. The movements of the planets are in the same direction, and almost in the same plane.

2. The movements of the satellites are in the same sense as those of the planets.

3. The rotations of these different bodies and the sun are in the same direction as their revolutions and in only slightly different planes.

4. The eccentricities of the orbits of the planets and their satellites are small.

5. Finally, the eccentricities of the orbits of comets are great, although their inclinations can be attributed to chance...*.

Whatever may be its nature, since it has produced or directed the movements of their satellites, this cause must have been common to all these bodies;

* Here Laplace interpolates a short discussion of Buffon's hypothesis, which will be dealt with in connexion with the tidal theories.

and seeing the enormous distance that separates them, it can have been only a fluid of immense extent. To have given them an almost circular motion around the sun in a single direction, the fluid must have surrounded the sun like an atmosphere. Consideration of the planetary movements leads us then to think that in consequence of an excessive heat the atmosphere of the sun formerly extended beyond the orbits of all the planets, and that it has since withdrawn to its actual limits. This might have taken place through causes similar to that which produced the brilliant outburst in 1572, lasting several months, of the famous star in the constellation Cassiopeia.

The great eccentricity of the orbits of comets leads to the same result. It evidently implies the disappearance of a large number of less eccentric orbits; which supposes an atmosphere around the sun extending beyond the perihelia of the observable comets, which, by destroying the movements of those that traversed it during its great extent, has reunited them to the sun. Thus we see that only the comets that were beyond in that interval should exist at the present time; and as we can observe only those which approach sufficiently near to the sun at perihelion, their orbits must be very eccentric. But we see at the same time that their inclinations must offer the same irregularities as if they had been projected at random, since the solar atmosphere has in no way affected their movements. Thus the long periods of revolution of comets, the great eccentricities of their orbits, and the variety of their inclinations, are very naturally explained by means of this atmosphere.

But how has it determined the revolutions and rotations of the planets? If these bodies had penetrated into the fluid, its resistance would have made them fall into the sun; it may then be conjectured that they were formed at the successive limits of this atmosphere, by the condensation of the zones that it must have abandoned in the plane of its equator while cooling and condensing towards the surface of that luminary, as has been seen in the preceding book. We may conjecture again that the satellites have been formed in a similar manner by the atmospheres of the planets. The five facts asserted above follow naturally from these hypotheses, to which the rings of Saturn add a new degree of plausibility. Finally, if in the zones abandoned in succession by the solar atmosphere, there were some molecules too volatile to unite with each other or to heavenly bodies, they must, while continuing to revolve around the sun, present to us all the appearances of the zodiacal light, without offering any noticeable resistance to the movements of the planets.

Whatever may become of this genesis of the planetary system, which I present with the lack of confidence that everything that is in no respect a result of observation or calculation must inspire, it is certain that its elements are so regulated that it must enjoy the greatest stability, so long as outside causes do not disturb it.

1·2. *The Dynamics of the Nebular Hypothesis.* The purely qualitative discussion given by Laplace is from its nature necessarily extremely incomplete, and in the absence of quantitative treatment it was, as he says, impossible to feel much confidence in the hypothesis. He never carried out such treatment, and for over sixty years the question of the mode of formation of the solar system remained as he had left it. Before proceeding to the later investigations, however, it is desirable to dwell further on Laplace's account, and to point out the places where his

development is incomplete and the modifications produced by modern observation in the empirical data he used.

The solar atmosphere whence the planets are supposed to have been formed was originally a highly diffuse nebula; this was apparently considered to have been approximately symmetrical about an axis nearly perpendicular to the plane of the ecliptic, since it is described as extending beyond the orbits of all the planets and as bearing some resemblance to Saturn's ring. The motion of the nebula is only stated to have been one of rotation; it is not made clear whether the periods of revolution of all parts of the nebula were to be the same, or whether, as might seem possible, the inner parts would revolve more quickly than the outer ones. The distribution of density is not definitely specified; but the description of the nebula as surrounding the sun like an atmosphere suggests that the sun was already in a fairly advanced state of condensation, and that the total mass of the nebula was probably less than that of the sun.

The nebula condensed slowly, and its rotation became faster and faster, in accordance with the dynamical principle of the conservation of angular momentum. Laplace assumes that his system is unaffected by outside disturbances. In such a system suppose the mass of some particle to be m, and consider the motion of the foot of the perpendicular from that particle to some fixed plane. Suppose this point to be moving with velocity v, and that the perpendicular from a fixed point in the plane to the tangent to the path is of length p. Then the principle asserts that if the values of mpv for all particles of the system are added up, the sum is constant for all time, no matter what changes occur within the system. This sum is called the angular momentum of the system about the given axis. Let us now apply this principle to the nebula. If r denotes the length of the perpendicular from a particle on to the axis of rotation at time t, while the fixed point and plane are the centre and the equatorial plane respectively, the sum is Σmrv, where as usual Σ is used to denote 'the sum of the values of.' Now suppose that the contraction takes place in such a way that the values of r for all parts of the nebula are altered in the same ratio. If a particle was initially at distance r_0 we shall then have

$$r = r_0 f(t),$$

where $f(t)$ depends on t alone. Suppose that the same holds for the velocity, so that if the initial velocity of a particle was v_0, we have

$$v = v_0 g(t).$$

As the mass of any particle remains constant, it follows from the above principle that

$$\Sigma m r_0 v_0 f(t)\, g(t) = \Sigma m r_0 v_0.$$

But since r_0 and v_0 do not depend on the time, we must have

$$f(t) = 1/g(t).$$

Now if the equatorial radius of the nebula is a, it is clear that a is pro-

portional to $f(t)$, and that the velocity at the outer boundary is proportional to $g(t)$ and therefore to $1/a$.)

The outer portions of the nebula at any time would be describing circular paths about the axis, under the gravitational attraction towards the interior and the fluid pressure acting outwards. The acceleration towards the axis required to maintain a circular path is v^2/r and is therefore proportional to a^{-3}. Since the action of fluid pressure is always outwards, rupture will take place if this acceleration is greater than that produced by gravity, namely a constant multiple of M/a^2, where M is the mass of the whole system. Hence as the mass contracts the acceleration necessary to hold it together increases like a^{-3}, while that available for the purpose only increases like a^{-2}. Therefore, whatever may have been the initial velocity of rotation of the mass, it is only a question of waiting till it has contracted sufficiently for gravity to become inadequate to hold the outermost portions in contact with the remainder. When this is so, they will be left behind. Rupture must take place at this stage unless it has previously taken place in some other way.

The foregoing argument is of considerable generality, for it covers the extreme cases where the mass rotates like a rigid body (that is to say, all parts take the same time to revolve around the axis) and either is homogeneous or has nearly the whole of the matter concentrated into a central nucleus. An indefinite number of possible intermediate states are also covered, and also a still wider range of states where the motion is not like that of a rigid body.

Given some rotation to start with, the advance of condensation will necessarily tend to produce some kind of rupture. It may happen that the nebula will become liquid or solid before conditions suitable for rupture are reached; condensation will then cease and rupture will be postponed indefinitely. A certain minimum angular momentum is therefore necessary to produce ultimate rupture in any given mass. But supposing such an amount of angular momentum to be available, rupture is certain to occur before the whole mass has been absorbed into the central body. It may take place at an earlier stage than is indicated by the theory just given. If this is so, it happens while the gravitational acceleration available at the outer boundary is adequate to retain the matter there. Hence it takes place, not by crossing the outer boundary, but by internal condensation.

Laplace does not consider the possibility of the formation of planets by internal condensation, but only the separation of the outer portions around the equator. The whole system is perfectly symmetrical in these circumstances, and therefore the detached matter would form a ring. As condensation proceeded, other rings would be formed, and after its separation each ring would revolve independently about the central body with the period appropriate to its distance. Several objections have been

made against this part of the theory. It was suggested* that the result would not be a series of thick rings, nine in number, but a very large number of extremely thin ones, thus tending to account for minor planets and meteors rather than major planets. Other writers have argued that the absence of cohesion would make the separation quite continuous, whereas the formation of separate rings requires it to have been intermittent. These objections were partially answered by Roche†, in his detailed discussion of the Nebular Hypothesis, the irregularity being attributed to intermittent cooling of the central body; but his explanation seems very artificial.

Granting that separate rings could have been formed, it remains to show how a ring could have condensed into a single planet. On this point neither Laplace nor any of his followers offers any clear account. Such condensation appears to be impossible, in consequence of certain results obtained by J. Clerk Maxwell in his *Essay on the stability of the motion of Saturn's Rings*, published in 1859. He shows that a ring of particles surrounding a central body will be stable if the number of the particles does not exceed a limit depending on the mass of the ring, and unstable if the number is greater than this limit. In a fluid ring the number of particles is effectively infinite, and therefore the motion of such a ring is necessarily unstable, and the ring will break up. Now any disturbance of the motion of such a ring can be represented as composed of a number of waves travelling around it. Each wave will remain steady, die down, or increase independently of all the others. Maxwell shows that the shorter the wave length the more rapidly will a disturbance increase, and therefore the method of rupture is that the fluid ring first breaks up into a very large number of detached bodies; the disturbances of somewhat longer wave lengths then assert themselves and cause the masses to run into one another and combine until the number of separate masses is so reduced that the condition for stability is satisfied. Thus the rupture of a ring would lead to the formation of a ring of planets in nearly the same orbit and of comparable size, which is not a state now represented in the solar system. If Jupiter were divided into 49 fragments revolving about the sun, it would satisfy Maxwell's criterion of stability, and hence when this stage had been reached further aggregation would cease. Thus the formation of the solar system by the breakdown of rings is impossible.

Even at this stage the questionable assumptions of the hypothesis are not exhausted. If the planets were formed by this process, the fact that they all revolve in one direction is explained; so are the small inclinations and eccentricities of their orbits. But when we come to the

* Daniel Kirkwood, *M.N.R.A.S.* **29**, 1869, 96–102.

† 'Mémoire sur la figure des atmosphères des corps célestes,' *Mémoires de l'Acad. de Montpellier*, **2**, 1854, 399; 'Essai sur la constitution et l'origine du système solaire,' *loc. cit.* **8**, 1873, 235. H. Poincaré, *Leçons sur les hypothèses cosmogoniques*, 1913, 15–67.

satellites there are further difficulties. Each planet is supposed to have acquired, by some unspecified process, a rotation in the same direction as its revolution, and the condensation of the planets is supposed to have afterwards led to the formation of systems of satellites by the same mechanism as produced the planets in the first place. Thus the fact, pointed out by Laplace, that most satellites revolve in the same sense as their primaries, is explained. But unfortunately for the theory eight satellites now known revolve in the opposite direction, a fact quite inexplicable by this means alone. It has been suggested that the rotations of the planets were originally retrograde, but were reversed by solar tidal friction during the condensation, but it has been found that this is inadequate to account for the direct rotation of Jupiter and Saturn (see 14·73).

Even in the case of the comets the theory is of doubtful value. The action of a resisting medium near perihelion would gradually reduce the mean distances of all comets that ever entered it, and therefore all surviving comets must have had perihelion distances greater than the present distance of Neptune. A further explanation of how these bodies came to be deflected so as to have small perihelion distances is therefore required, and is not provided by the theory.

1·3. *The Nebular Hypothesis with Internal Condensation.* Although the particular course of evolution of the solar nebula sketched out by Laplace has proved to be impossible, it does not follow that an extended nebula could not develop into a system very like the solar system by some different process. It is noticed on examining Laplace's discussion that the abandonment of perfect symmetry takes place at the very last moment conceivable, and that so long as we admit only symmetrical forms the only course possible is one such that the nebula sheds matter continuously around its equator, thus forming a new nebula around itself with a different distribution of velocity and probably of density. Now to produce the planetary system, symmetry must be abandoned at some stage, and it has been shown that if the abandonment is postponed till rings have been formed it is by that time too late. Emission from the boundary is not possible in any other way; the only alternative left is by internal condensation, a possibility that has already been suggested. Such a theory would require that some slight inequality of density should be formed within the nebula, and that on account of a type of instability the inequality should proceed to increase until the form of the nebula was very seriously altered and one or more planetary nuclei formed. We see at once that this accounts for several features of the system, if it is a possible mode of development. The direct revolution of the planets in slightly inclined and slightly eccentric orbits is explained. The rotation of a planet and the revolution of its satellites would be in the direction of rotation of the matter in the

neighbourhood of the primitive condensation, and would be direct if the condensation started within the inner portion, but perhaps reversed if within the matter already shed from around the equator. Thus it might be possible to explain why the most remote planets have retrograde satellites; but there appears to be still no way of explaining the existence of both direct and retrograde satellites of the same planet, as is true of both Jupiter and Saturn.

1·31. This hypothesis, however, fails on other grounds. The principle of the constancy of angular momentum, used at the very outset, imposes a severe restriction on the initial conditions of the mass, which turns out to be inconsistent with the condition that condensation shall be possible at all. If we use astronomical units of length, time, and mass, so that the earth's distance from the sun and the sun's mass have measure unity and the year has measure 2π, we can calculate the angular momentum of the whole system at present about an axis through the centre of the sun perpendicular to the ecliptic, and this cannot have changed since the earliest times. The data for the various planets are as follows:

Body	Mass $1\div$	Mean distance	Mean motion	Angular momentum
Sun	1	$2{\cdot}7 \times 10^{-3}$	14·4	$1{\cdot}05 \times 10^{-4}$
Mercury	9700 000	0·39	4·17	$6{\cdot}5 \times 10^{-8}$
Venus	402 000	0·72	1·63	$2{\cdot}1 \times 10^{-6}$
Earth	329 000	1·00	1·00	$3{\cdot}0 \times 10^{-6}$
Mars	3100 000	1·52	0·53	$4{\cdot}0 \times 10^{-7}$
Jupiter	1047	5·20	0·0844	$2{\cdot}18 \times 10^{-3}$
Saturn	3510	9·54	0·0367	$0{\cdot}95 \times 10^{-3}$
Uranus	22800	19·2	0·0119	$0{\cdot}19 \times 10^{-3}$
Neptune	19700	30·1	0·0061	$0{\cdot}28 \times 10^{-3}$

In this table the mean motion n is the mean angular velocity of the body as seen from the sun; it is also the reciprocal of the number of years in the period of revolution. The velocity is therefore rn, and the angular momentum is mr^2n. For the sun the mean distance given is k, the radius of gyration, defined as the distance from the axis of a particle with the whole mass of the sun that would have the proper angular momentum if it revolved in the sun's period of rotation. It is supposed that the density of the sun is uniform, so that

$$k^2 = \tfrac{2}{5}r_0^2 \qquad \text{.........(1),}$$

where r_0 is the radius of the sun. If the sun is condensed towards the centre this value will be reduced. The mean motion given for the sun is its angular velocity of rotation.

When we add up all the angular momenta, the total is found to be $3{\cdot}7 \times 10^{-3}$, of which Jupiter contributes more than half and the sun only about a thirtieth. The contributions of the inner planets, the satellites and the rotations of the great planets on their axes are all inappreciable. Now if the sun and all the planets were united into a homogeneous sphere

or ellipsoid filling the orbit of Neptune and rotating in Neptune's actual period, its angular momentum would be $\frac{2}{5}Mr_9{}^2n_9$, where M is the total mass of the sun and planets and r_9 and n_9 refer to the mean distance and motion of Neptune. This amounts to about 2·2, about 600 times the actual angular momentum of the system.

Now a condensation within the mass must have revolved in the same period as the mass itself, and thus the angular velocity of the mass when Neptune was formed must have been n_9. Hence, if k_9 be the radius of gyration at that time, we must have

$$Mk_9{}^2n_9 = 3{\cdot}7 \times 10^{-3} \qquad \text{.........(2)},$$

$$\tfrac{2}{5}Mr_9{}^2n_9 = 2{\cdot}2 \qquad \text{.........(3)},$$

whence

$$k_9{}^2/\tfrac{2}{5}r_9{}^2 = 1/600 \qquad \text{.........(4)}.$$

Thus the radius of gyration of the nebula was only $\frac{1}{26}$ of that of a homogeneous ellipsoid of the same diameter, indicating that its mass was strongly condensed towards the centre, even in the early stages. The suggestion developed by Roche (*loc. cit.*) that the sun was already formed when the condensation started is therefore essential to the theory*.

1·32. The effect of this central condensation is that the gravitation of the mass reduces practically to that of a heavy particle at the centre, and in these circumstances it is pointed out by Jeans† that there are no possible asymmetrical steady states at all; the symmetrical state is always stable, and the course of development will therefore proceed through all the symmetrical forms found by Roche until ejection commences around the equator. Hence internal condensations are impossible and the theory fails.

1·33. The argument may be presented in another way by means of a theorem due to Poincaré. If the angular velocity of the fluid in the neighbourhood of the suggested condensation is ω, consider the function

$$U_1 = U + \tfrac{1}{2}\omega^2(x^2 + y^2) \qquad \text{.........(5)},$$

where U is the gravitational potential and x and y are rectangular coordinates measured from some point within the condensation and parallel to the equatorial plane. Consider any closed surface within the condensation. Then we have by Green's theorem

$$\iint \frac{\partial U_1}{\partial n}\,dS = \iiint \nabla^2 U_1\,d\tau \qquad \text{.........(6)},$$

* The importance of the angular momentum criterion was first pointed out by Babinet (*Comptes Rendus*, **52**, 1861, 481–84) and by David Trowbridge (*Amer. Journ. of Science and Arts*, Ser. 2, **38**, 1864, 344–60). Both of these writers, however, overlooked the fact that the orbital angular momentum of Jupiter much exceeds the rotational angular momentum of the sun, and attended only to the latter. The first correct discussion is by Fouché (*Comptes Rendus*, **99**, 1884, 903–906). For an account of the primitive density distribution necessary to account for all the planets, reference may be made to Anne Sewell Young (*Astrophysical Journal*, **13**, 1901, 338–43), who undertook the work at the suggestion of Prof. F. R. Moulton.

† *Problems of Cosmogony and Stellar Dynamics*, p. 148.

the integral on the left being taken over the surface and that on the right throughout the interior. Now

$$\nabla^2 U_1 = \nabla^2 U + 2\omega^2$$
$$= -4\pi f\rho + 2\omega^2 \qquad \text{.........(7)},$$

where ρ is the density and f the constant of gravitation. If ρ_1 is the mean density inside this surface and ϕ its volume, the integral on the right is therefore equal to $2\phi(\omega^2 - 2\pi f\rho_1)$.

Now $-\dfrac{\partial U}{\partial n}$ is the normal component of gravity inwards across the surface, and $\dfrac{\partial}{\partial n}\tfrac{1}{2}\omega^2(x^2+y^2)$ is the inward normal acceleration required to keep a particle in contact with a surface rotating with angular velocity ω. Hence if

$$\frac{\partial U_1}{\partial n} = \frac{\partial U}{\partial n} + \frac{\partial}{\partial n}\tfrac{1}{2}\omega^2(x^2+y^2) \qquad \text{.........(8)}$$

is positive, the inward attraction due to gravity is less than is required to hold a particle in contact with the surface, and thus the combined effect of gravity and rotation will tend to make the fluid move outwards across the closed surface. In addition the fluid pressure must increase inwards, since the pressure must increase with the density, and therefore the effect of pressure is always to produce motion outwards. But for a condensation to develop it is necessary that the fluid shall have at all points an inward or a zero acceleration, and therefore $\dfrac{\partial U_1}{\partial n}$ must be negative and sufficiently great to counteract the outward acceleration due to pressure. This must be true at all points of the surface if the effect is to be a permanent condensation around a point. Hence $\displaystyle\iint \frac{\partial U_1}{\partial n}\, dS$ over the surface must be negative, and therefore $2\phi(\omega^2 - 2\pi f\rho_1)$ must be negative. Thus the condition for condensation to be possible is that $2\pi f\rho_1$ shall be greater than ω^2.

If the whole mass rotated like a rigid body, ω would be the same at all points, and therefore equal to n, the angular velocity of the whole nebula, and to the present mean motion of the planet formed at that stage. Now by Kepler's third law we have

$$\omega^2 = \frac{fM}{r^3} \qquad \text{.........(9)}.$$

Hence the condition for condensation is that $2\pi f\rho_1$ shall be greater than fM/r^3, or that ρ_1 shall be greater than $M/2\pi r^3$. But the angular momentum condition (4) shows that the density near the outer boundary is only a small fraction of the mean, and in this case the form of the outer boundary is such that

$$M/2\pi r^3 = 0{\cdot}36\bar{\rho} \qquad \text{.........(10)},$$

where $\bar{\rho}$ is the mean density*. Hence the condition for condensation is

* Jeans, *loc. cit.* p. 150.

that the density in the neighbourhood of the condensation when condensation begins shall be greater than 0·36 of the mean density*; which contradicts the result that nearly all the mass must be near the centre, the density as far out as the orbit of Neptune being at most only of the order of a thousandth of the mean.

1·34. If we suppose condensation to have occurred in the matter that has been shed around the equator, a similar contradiction arises. Each portion of the matter in this part must revolve independently about the centre, with an angular velocity nearly equal to n, that appropriate to a circular orbit at the same distance, namely $(fM/r^3)^{\frac{1}{2}}$. Now the rotation of the matter is given by

$$2\omega = \frac{1}{r}\frac{d}{dr}(r^2 n) = \tfrac{1}{2} n \qquad \text{.........(1).}$$

Thus the condition for condensation is that ρ_1 shall be greater than $\omega^2/2\pi f$ or $n^2/32\pi f$, and finally greater than $0{\cdot}022\bar{\rho}$—much less than in the former case, but still quite inconsistent with the data of the problem. Hence internal condensation is impossible in either portion of the nebula.

1·4. *Summary.* The theory of rotational instability is therefore not a possible explanation of the origin of the solar system, and search must be made elsewhere for the correct theory. It has actually been proved by Jeans in his elaborate investigation that a nearly homogeneous mass broken up by rotational instability would give rise to a double or multiple star, the masses of the components being comparable; while a mass with a strong central condensation, if it condensed elsewhere at all, would probably give a spiral nebula, the arms consisting of streams of stars, each with a mass comparable with that of the sun. In neither case would anything resembling the solar system be produced. A gaseous body with a mass comparable with that of the sun, and strongly condensed towards the centre, could not condense at all except by absorption into the sun.

* A very similar criterion, on rather different hypotheses, was given by Moulton, *Astrophysical Journal*, 11, 1900, 122–6. It forms part of a paper giving the objections to Laplace's theory found by Moulton and others up to that date.

CHAPTER II

The Tidal Theory of the Origin of the Solar System

"In six days the Lord made heaven and earth, the sea, and all that in them is." Exodus xx. 11.

2·1. Long before it had been definitely proved that Laplace's Nebular Hypothesis could not give a dynamically satisfactory account of the origin of the solar system, and could not be modified to give one so long as the notion of gradual development from a nebula initially fairly symmetrical was retained, many astronomers had felt that it involved too many unproved and unplausible assumptions, and were seeking for an alternative hypothesis that would avoid the difficulties. The chief of these alternatives was the Planetesimal Hypothesis of Professors T. C. Chamberlin and F. R. Moulton, both of Chicago*; this hypothesis has obtained the support of a considerable number of geologists, although astronomers have given it much less attention. Nevertheless its astronomical side appears to contain a large amount of truth, in spite of several serious objections that can be urged against it. One of these objections was pointed out by me in 1916†, but the advantages of this hypothesis over those previously advocated were so striking that I attempted to modify it in such a way as to avoid the objection. An outline of the result was published in 1917‡, and various applications of the modified hypothesis to account for phenomena of the existing solar system were developed in 1918§. One of the chief changes found necessary was to abandon the part of the theory that attributed any cosmogonic importance to the planetesimals, and the modified theory is therefore known as the Tidal Theory and not as the Planetesimal Theory.

2·2. *The Methods of Rupture.* Sir J. H. Jeans has discussed on general dynamical principles the whole problem of the possible methods of rupture of fluid masses, and was able to show, in a thesis|| awarded the Adams Prize of the University of Cambridge in 1917, that the only method that could lead to anything resembling the solar system required almost the same initial conditions as the Planetesimal Hypothesis and that which

* Chamberlin, *Astrophysical Journal*, **14**, 1901, 17–40; Moulton, *loc. cit.* **22**, 1905, 165–81; also references in Appendix A.

† *M.N.R.A.S.* **77**, 1916, 84–112.

‡ *Science Progress*, No. 45, 1917, 52–62.

§ *M.N.R.A.S.* **78**, 1918, 424–41.

|| *Problems of Cosmogony and Stellar Dynamics*, 1919; also *Memoirs R.A.S.* **62**, 1917, 1–48.

I adopted. Since the two discussions rested on very different data, the close agreement between the inferences is a strong argument for the truth of some closely similar theory. There were, however, a few points of difference; these will be indicated in the following account, which differs somewhat from both presentations. For convenience of treatment, the discussion of the Planetesimal Hypothesis itself will be reserved until Appendix A.

The fundamental feature of the hypothesis is the approach to the sun of a star considerably more massive than itself. This raised two large tides on the sun, the greatest protuberances being at the points of the sun nearest to and furthest from the star. When the distance between the two bodies became sufficiently small, the tendency to disruption due to the difference between the attractions of the star on the two opposite sides of the sun became greater than the sun's gravitation could counteract, and a portion of the sun was torn away. This afterwards condensed to form the planets and satellites.

Jeans has shown that in the tidal theory, as in the rotational theory, a mass may be broken up in two ways, according as it is approximately homogeneous or strongly condensed towards the centre. The former type would give a number of bodies with masses of the same order of magnitude; the latter is the type required to account for the solar system, and will therefore be considered in some detail.

The gravitation potential due to the sun and a star together, when the sun is supposed to have nearly all its mass concentrated near the centre and the star is supposed spherical, is

$$U = \frac{fM}{r} + \frac{fM'}{r'} \qquad \text{.........(1)},$$

where f is the gravitational constant, M and M' are the masses of the sun and star, and r and r' the distances of the point considered from the centres of the respective bodies. If coordinates x, y and z be taken, the origin being at the centre of the sun and the x axis pointing towards the star, and if u, v, w be the component velocities of the matter of the solar envelope at any point, we have

$$\frac{d}{dt}(u, v, w) = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right) U - \frac{1}{\rho}\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right) p \qquad \text{.........(2)},$$

where the derivatives with regard to the time indicate the accelerations in the directions of the axes, p is the pressure and ρ the density. Now the acceleration of the central body towards the star will be fM'/R^2, where R is the distance between the centres. Hence if (u_0, v_0, w_0) be the velocity of the centre of the sun,

$$\frac{d}{dt}(u_0, v_0, w_0) = \left(\frac{fM'}{R^2}, 0, 0\right) \qquad \text{.........(3)},$$

and if (u', v', w') be the velocity of a particle of the envelope relative to the centre of the sun, we have, since R is independent of x, y and z,

$$\begin{aligned}\frac{d}{dt}(u', v', w') &= \frac{d}{dt}(u - u_0, v - v_0, w - w_0)\\ &= \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right)\left(U - \frac{fM'x}{R^2}\right)\\ &\quad - \frac{1}{\rho}\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right)p \qquad \text{.........(4).}\end{aligned}$$

The discussion of the resulting motion may be facilitated by considering two extreme cases, called by Jeans 'slow' and 'transitory' encounters. In a slow encounter, the changes are slow compared with the corresponding free vibration of the solar envelope, while in a transitory one they are quick.

The changes during a slow encounter bear a certain resemblance to the motion of a pendulum when the bob is drawn aside by a gradually increasing horizontal force. It moves slowly in the direction of the force, the deflexion at any instant being almost the same as if the pendulum were in equilibrium under the action of gravity and the force actually acting at that instant. In a slow tidal encounter, approximate equilibrium is similarly maintained, and the form of the solar envelope can be calculated as if it were in equilibrium under the action of the gravitation of the sun and the star. Thus the acceleration $\frac{d}{dt}(u', v', w')$ can be neglected.

In the case of the transitory encounter, the analogous pendulum problem is that of a bob suddenly struck by a horizontal blow, and then allowed to swing or revolve freely. In the tidal problem, the motion is as if a velocity was suddenly imparted to the envelope, and this was then left to readjust itself under the mutual attraction of its parts and the sun's gravitational field.

2·21. In a slow encounter we have

$$\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right)\left(U - \frac{fM'x}{R^2}\right) = \frac{1}{\rho}\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right)p \qquad \text{.........(1).}$$

Thus the fluid pressure and the density must be functions of $U - \dfrac{fM'x}{R^2}$, that is, of $\dfrac{M}{r} + \dfrac{M'}{r'} - \dfrac{M'x}{R^2}$. Jeans (*loc. cit.* pp. 153–56) denotes this latter quantity by Ω. Thus the outer boundary of the mass will be one of the surfaces where Ω is constant; it will be the particular surface of the set that is just large enough in volume to be filled by the matter of the solar envelope. Jeans shows that one of these surfaces, represented by the thick curve in Fig. 1, surrounds the sun completely and has a sharp point on the side nearest the star. At first the volume of the solar envelope is much less than that of this surface; but as the star approaches, the linear dimensions of this critical surface diminish in proportion to the

distance between the two bodies, until the volume becomes too small to hold the envelope. The critical value of R, if the disturbing body has a mass double that of the sun, is $2{\cdot}87a$, where a is the undisturbed radius of the envelope. When this state is reached $\frac{\partial \Omega}{\partial x}$ vanishes at the apex; and this implies that whereas up to this stage everywhere, and even at this stage everywhere else, the resultant influence of gravity is to accelerate the matter towards the interior, the influence now ceases at this point. With a closer approach it changes its direction, and the matter of the envelope is drawn out at the point in the direction of the star. The action of fluid pressure is always to encourage ejection, which will therefore continue until the star has receded again to such a distance that its gravity is no longer enough to neutralize that of the sun.

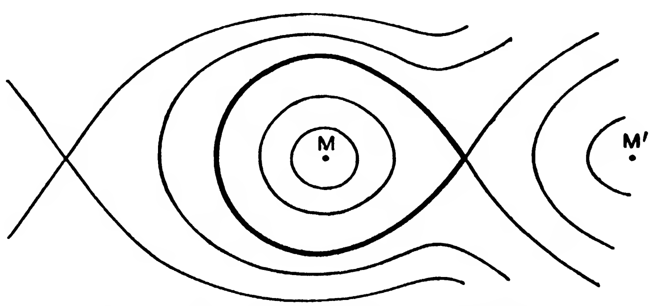


Fig. 1. (From Jeans's *Problems of Cosmogony.*)

Initially the velocity and acceleration of the ejected matter relative to the sun are both inappreciable, but when the star approaches closer the acceleration increases and the matter therefore begins to move straight towards the star. Meanwhile the star is moving transversely as well as towards the sun; thus, when the matter has moved some distance outwards, the star is no longer in its direction of motion, and will therefore attract it sideways. Thus a velocity around the sun as well as away from it will be acquired. The changes in the condition of the system are indicated in the following diagram (p. 20), where the dotted curve shows the position of the solar envelope at the first rupture and the figures 1, 2, and 3 indicate the positions and paths of the particles ejected at the first, intermediate, and final stages. Since all the accelerations relative to the sun are in one plane, namely that of the motion of the star relative to the sun, it follows that the whole of the motion produced must be approximately in this plane, apart from a possible slight departure due to the rotation of the sun before the passage of the star. Further, the transverse motion of every portion of the ejected matter is in the same sense as the motion of the star. It would be incorrect, however, to suppose that the whole of the ejected

matter would follow, even at first, the same orbit. The point where emission is taking place is necessarily immediately towards the star, and therefore is continually changing as the star moves, its path relative to the centre of the sun being geometrically similar to the relative path of the star. Thus all particles start from different positions and therefore travel on different paths; they do not follow one another along one path, as might at first be thought.

If ejection was continuous, all the ejected matter would at any subsequent time lie within a long narrow region; in other words, it would form a filament. In shape this would resemble a boomerang. For, subject to the conditions of a slow encounter, the velocity of the matter relative to the sun is inappreciable before ejection, and the acceleration is zero at the moment of ejection. Hence the end of the filament nearest the centre must touch the locus of the equilibrium point, while the acceleration away from the sun as the star approaches will produce a considerable outward velocity, so that the filament will be curved away from the sun.

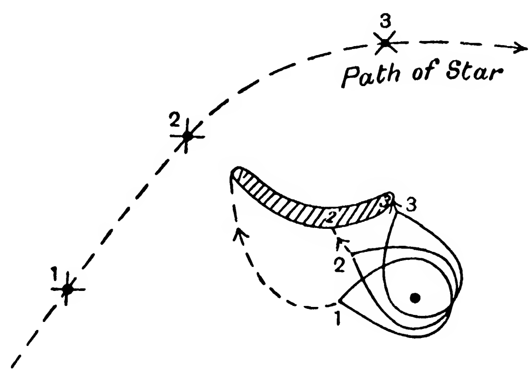


Fig. 2. Diagram of the changes in the form of the solar envelope and the paths relative to the sun of portions of the ejected matter during the passage of the star. The position of the filament corresponds to position 3 of the star.

It will be seen that when the star reaches its nearest point to the sun, the point of ejection is also at its nearest. When the star recedes in the least, the sun's attraction on the matter severed at the moment of closest approach will again be greater than the tidal disturbing force, and it will therefore draw this matter back into the solar envelope. We see similarly that no ejection of matter not already separated can take place during the retreat of the star, and indeed that what has been ejected during the approach will fall back, with the exception of such as has acquired a sufficient transverse velocity to miss the envelope when it approaches the sun on its next revolution. Thus only the parts first drawn off can remain permanently detached from the sun. At any stage of

ejection the volume of the critical surface would be decreasing at a rate simply proportional to the rate of change of volume of a sphere with its centre at the sun and extending to the star. Thus when ejection started the rate of ejection would be finite if the rate of approach was finite, would increase to a maximum, and would fall to zero at periastron*.

The typical slow encounter just described is, however, incapable of being realized in actuality, for quantities supposed negligible in it are in fact incapable of being small. The period of a free vibration of tidal type in the solar envelope could hardly differ in order of magnitude from the period of revolution of a planet moving in a circular orbit at the point of ejection. This period is $2\pi\,(b^3/fM)^{\frac{1}{2}}$, where b is the radius to the point of ejection. The time of passage of the star, even if the relative velocity before it was appreciably affected by the sun was inappreciable, would be of the order of $\pi\,\{R^3/f\,(M+M')\}^{\frac{1}{2}}$, by the ordinary theory of two bodies. Now the condition for any emission to occur is approximately that b shall exceed $(M/2M')^{\frac{1}{3}}R$. Thus the ratio of the two times considered is of order $\left(\dfrac{2\,(M+M')}{M'}\right)^{\frac{1}{2}}$, which cannot be less than unity, whereas it would have to be a small fraction for the conditions of a slow encounter to be satisfied. Any actual encounter would be intermediate in character between the typical slow encounter and the typical transitory one; but whereas the former type cannot be even approximately realized, being dynamically impossible for freely moving bodies, the latter can be realized with any required degree of accuracy.

2·22. Consider now a transitory encounter. The equation of continuity is

$$-\frac{1}{\rho}\frac{d\rho}{dt}=\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}+\frac{\partial w}{\partial z}\qquad\text{.........(1)},$$

where the axes are now taken to be fixed in direction and not accelerated, and u, v, w are the component velocities of the fluid. Now by hypothesis the encounter is such as will make the velocities moderate†, while its duration is inappreciable. Hence the quantity on the right is moderate, and therefore its integral for the whole or any part of the encounter is inappreciable; thus ρ, the density of any element of fluid, does not change during the encounter.

* This account differs from that of Jeans, *loc. cit.* p. 283: "The rate of ejection of matter would be slow at first, would increase to a maximum when the passing star was at its nearest to the sun and would subsequently diminish to zero."

† In arguments based chiefly upon the notion of orders of magnitude, some set of quantities must be taken as a standard. Other quantities are called 'large,' 'small' or 'moderate' in comparison with these. The standard quantities must, of course, be consistent in the same sense as is applied to units of measurement. Thus, in the present problem the mass of the sun, gravity at the surface of the primitive sun, the velocity and mean motion of a planet at the sun's surface and moving about it in a circular orbit, and the radius of the sun, are all standard quantities, while the pressure at the sun's centre is moderate.

If the fluid be compressible, the pressure on any element of the fluid is determinate when the density is known, and therefore, by the last paragraph, does not change during the encounter. Now the accelerations of any element are given by

$$\frac{d}{dt}(u, v, w) = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right) U - \frac{1}{\rho}\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right) p \quad \text{.........(2).}$$

Let us denote $\int U dt$, taken for the whole duration of the encounter, by Φ. By the last paragraph, the second vector on the right is essentially moderate, and its integral throughout the encounter is therefore negligible. Thus if we integrate the equations of motion for the duration of the encounter we shall have

$$u, v, w = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right) \Phi \quad \text{.........(3).}$$

Thus there is no impulsive pressure during a transitory encounter.

If the fluid be incompressible, the argument breaks down, but the result may be obtained otherwise. Equations (2) still hold, but, ρ being constant for any given element, we can denote $\int p dt$ for the duration of the encounter by ϖ, and then if we integrate (2) for the duration, we have

$$(u, v, w) = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right) \Phi - \frac{1}{\rho}\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right) \varpi \quad \text{.........(4).}$$

Differentiating the three equations thus combined with regard to x, y, and z respectively, and adding, we have

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = \nabla^2\Phi - \left\{\frac{\partial}{\partial x}\left(\frac{1}{\rho}\frac{\partial \varpi}{\partial x}\right) + \frac{\partial}{\partial y}\left(\frac{1}{\rho}\frac{\partial \varpi}{\partial y}\right) + \frac{\partial}{\partial z}\left(\frac{1}{\rho}\frac{\partial \varpi}{\partial z}\right)\right\} \quad \text{...(5).}$$

Now the left side is equal to $-\frac{1}{\rho}\frac{d\rho}{dt}$, which is zero since the fluid is incompressible. $\nabla^2 U$ is equal to $-4\pi f\rho$, which is always moderate, and therefore $\nabla^2\Phi$ is inappreciable. Hence

$$\frac{\partial}{\partial x}\left(\frac{1}{\rho}\frac{\partial \varpi}{\partial x}\right) + \frac{\partial}{\partial y}\left(\frac{1}{\rho}\frac{\partial \varpi}{\partial y}\right) + \frac{\partial}{\partial z}\left(\frac{1}{\rho}\frac{\partial \varpi}{\partial z}\right) = 0 \quad \text{.........(6).}$$

If we had an uncharged dielectric with a specific inductive capacity proportional to $\frac{1}{\rho}$, this differential equation would be satisfied if ϖ were the electrostatic potential within it. Further, the pressure, and therefore ϖ, vanish at all points of the boundary of the envelope. The problem is therefore completely analogous to that of an uncharged dielectric with a zero potential over the boundary, which is known to have a unique solution*, namely that the potential is zero for all points within it. Hence we have the general result that there is no impulsive pressure in a transitory encounter, and therefore equations (3) hold in all cases.

* Cf. Jeans, *Electricity and Magnetism*, 1908, 161–162.

2·221. Now by hypothesis the gravitation of the sun is inappreciable in comparison with that of the star during the encounter, even within its envelope, and *a fortiori* it is therefore insufficient to divert the path of the star appreciably. The star will therefore move in a straight line. Taking the path of the star to be the axis of y and its velocity to be v, the coordinates of the star are $(0, vt, 0)$ and the accelerations of a particle (x, y, z) due to the star are

$$-fM'\left(\frac{x}{r'^3}, \frac{y-vt}{r'^3}, \frac{z}{r'^3}\right),$$

where r' is the distance of the particle from the star.

Integration of these with regard to the time gives the velocity of the particle at the end of the encounter. Putting $x^2 + z^2$ equal to R_1^2, so that R_1 is the least distance of the star from the particle considered, and $y - vt$ equal to $R_1 \tan\phi$, we have

$$r'^2 = x^2 + z^2 + (y - vt)^2$$
$$= R_1^2 \sec^2\phi,$$

and $$-vdt = R_1 \sec^2\phi\, d\phi.$$

The direction of the final velocity evidently intersects the path of the star. The y component is

$$\frac{fM'}{R_1 v}\int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} \sin\phi\, d\phi,$$

which vanishes. The component at right angles to the path of the star is

$$\int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} \frac{fM'}{R_1^3 \sec^3\phi} . R_1 . \frac{R_1}{v} \sec^2\phi\, d\phi = \frac{2fM'}{R_1 v}.$$

If R_0 be the distance of closest approach to the centre of the sun, and if

$$R_1 = R_0 - a,$$

the velocity of the particle relative to the centre of the sun at the end of the encounter will be

$$\frac{2fM'}{v}\left(\frac{1}{R_0 - a} - \frac{1}{R_0}\right) = \frac{2fM'a}{R_0^2 v} \quad \text{.........(1)}$$

approximately.

The condition that the sun may be broken up is evidently that this velocity shall be comparable with the parabolic velocity for a particle at the outside of the envelope. If a be the mean radius, this condition becomes

$$\frac{2fM'a}{R_0^2 v} > \left(\frac{2fM}{a}\right)^{\frac{1}{2}} \quad \text{.........(2).}$$

We have the further condition that the duration of the encounter shall be less than the period of oscillation of the mass. The former can be taken

to be $2R_0/v$, and the latter half the period of revolution of a planet at the boundary. Hence

$$\pi\left(\frac{fM}{a^3}\right)^{-\frac{1}{2}} > \frac{2R_0}{v} \qquad \text{.........(3).}$$

Multiplying (2) and (3), we find on simplification

$$\frac{\pi}{\sqrt{2}}\frac{M'}{M} > \frac{R_0{}^3}{a^3} \qquad \text{.........(4).}$$

The corresponding condition found for a slow encounter was equivalent to

$$\frac{2M'}{M} > \frac{R_0{}^3}{a^3} \qquad \text{.........(5).}$$

Thus for a given periastron distance the least mass of the disturbing body that will produce rupture is found to be nearly the same whether the encounter is slow or transitory. But we notice from (2) that if a and R_0 are fixed the critical value of M' will increase with v.

For a given mass M' and a given minimum separation of the two bodies, equation (2) imposes an upper limit to the values of v consistent with rupture taking place. To indicate the possible dimensions of the system at rupture, take

$$M' = 4M,\ R = 1{\cdot}5a,\ v = 4 \times 10^6 \text{ cm./sec.}$$

With $\qquad M = 2 \times 10^{33}\text{ gm.},\ f = 6{\cdot}6 \times 10^{-8}\text{ c.g.s.},$

(2) gives $\qquad a < 6{\cdot}4 \times 10^{13}\text{ cm.},$

which is the present mean distance of a rather remote asteroid.

The extreme type of transitory encounter may be approached indefinitely closely by supposing M' and v increased indefinitely while retaining a constant ratio to each other. It is easy to see, however, that this type of encounter could not lead to anything resembling the solar system. The investigation just given shows that in a purely transitory encounter the motion produced necessarily consists of the shooting out of two protuberances, one towards the star and one away from it, but both directly away from the centre of the sun. Some of the matter might acquire enough velocity to make it leave the sun's influence altogether; but such as did not would fall straight back into the sun and be reabsorbed. Thus no planetary system would be generated. To account for the actual solar system an encounter of intermediate type is required.

It has been suggested at various times that if the solar system was formed in this way, the sun might at the same time have broken up the star, which would thereby have acquired a set of planets of its own. This, however, is not possible. Equations (4) or (5) can be written

$$\frac{M'}{R_0{}^3} > \frac{M}{2a^3}.$$

This is the condition that the mass M shall be broken up. The corresponding condition that M' shall be broken up is

$$\frac{M}{R_0^3} > \frac{M'}{2a'^3} \qquad \text{.........(6)},$$

where a' is the radius of M'. From these two equations we get

$$R_0^2 < 2^{\frac{2}{3}} aa' \qquad \text{.........(7)}.$$

But $$R_0^2 > (a + a')^2 > 4aa' \qquad \text{.........(8)},$$

and therefore (7) cannot be satisfied. Only one of the stars concerned in a tidal encounter can be broken up.

2·3. *The Density of the Primitive Sun.* So far the only postulate we have made about the condition of the primitive sun is that its density was not nearly uniform; it must have been strongly condensed towards the centre. There is good reason, however, to believe* that the time that has elapsed since the events we are discussing is so small a fraction of the life of a star that we are justified in supposing the sun to have been in practically its present state. The present mass, radius, and mean density of the sun will therefore be adopted here.

2·4. *The Rupture of the Filament.* Whether the encounter was slow or transitory, the result must therefore have been the formation of a long protuberance towards the star; transverse motion, if any, would be in the plane of the star's relative motion. In either case it is possible, though not necessary, for a shorter protuberance to be formed on the side away from the star; this depends on the mass of the star and on its distance of closest approach. In the slow encounter the filament would be formed by ejection of matter from a limited region of the boundary; the shorter filament, if formed, would be diametrically opposite to it. In a transitory encounter the whole envelope would commence to be stretched, the outward velocity relative to the central body being greatest on the side nearest to the periastron, a smaller outward velocity being also produced on the opposite side. The actual encounter would be of intermediate type. The longer protuberance produced will in general be called 'the filament.'

We have now to consider the subsequent development of the system. As the star retreated, much of the matter drawn out, including perhaps the whole of the shorter protuberance, must have fallen back into the sun. The outer portions of the longer one would, however, continue to recede from the sun, and, being deflected transversely by the star, would miss the sun on the return journey and continue to revolve around it.

* See later, p. 77.

There is no necessary upper limit to the length of a filament produced in this way; if the initial velocity was greater than the parabolic velocity at the point of emission, gravity would be unable to prevent the first matter drawn out from receding to an infinite distance, and the fluid pressure of the matter behind it would always be pushing it onwards. Thus we should expect a long narrow filament to be formed.

It may be mentioned that none of the ejected matter would be carried off by the star. This could happen only if this matter had *relative to the star* a velocity less than the parabolic velocity. But the velocity of the sun relative to the star was at all instants greater than the parabolic velocity, otherwise the sun and star would still be associated; and the ejected matter was initially stationary with reference to the sun. Hence its velocity relative to the star was always greater than the parabolic velocity. Some matter may have been lost to both bodies, and probably was; but none was carried off by the star.

If such a strip were steadily drawn out under the joint influence of the sun and the star, the mass per unit length in any region of it would presumably vary steadily with the distance from the sun, having possibly one maximum in the middle. If, however, it received a slight distortion, the mass per unit length in some region being increased, it is possible that the extra gravitative power of this region would draw other matter towards it. The disturbance would then increase exponentially with the time, and therefore a condensation would be formed in the filament.

The theory of the aggregation of a gaseous filament in this way has been outlined by Jeans (*loc. cit.* pp. 157–60). Condensation would begin when the length of the strip already ejected reached the value

$$l = \tfrac{1}{2}c\left(\frac{\pi}{f\rho}\right)^{\frac{1}{2}} \qquad \qquad \text{.........(1)},$$

where c is the velocity of sound in the gas. The ejected portion would then begin to detach itself from the main body; thenceforth it would lead an independent existence as a planet. If ejection continued, a second planet would be formed, and others would follow until ejection ceased. Those formed last would fall back into the sun as the disturbing body receded, but the early ones would retain their identity.

The above condition is necessary for condensation, but not sufficient. Two factors will tend to make a gaseous filament dissipate itself instead of condensing; namely, first, the tendency of a gas to spread itself out into the surrounding vacuum, and second, the tidal disruptive action of the sun itself. A gas suddenly liberated into a vacuum will spread out with a velocity equal to a moderate multiple of the velocity of sound in the gas. According to the usual, but probably inaccurate, formula*, the velocity

* Ramsey, *Hydromechanics*, Part II, 1913, 59.

of efflux of a gas into a vacuum should be $\left(\frac{2}{\gamma - 1}\right)^{\frac{1}{2}} c$, where γ is the ratio of the specific heats. For a monatomic gas this amounts to $\sqrt{3}c$. The condition that the filament should not dissipate itself is that this must be less than the velocity necessary to remove a particle from the gravitative influence of the filament. It is, however, more convenient mathematically to consider the corresponding criterion for a primitive planet. If a be the radius of such a planet, and m its mass, this condition shows that

$$3c^2 < 2fm/a \qquad \text{.........(2).}$$

If b be the radius of the filament and we assume that the density did not change considerably while a section was collecting into a nearly spherical form, we have

$$m = \pi\rho b^2 l = \tfrac{4}{3}\pi\rho a^3 \qquad \text{.........(3),}$$

and therefore

$$3c^2 < \tfrac{8}{3}\pi f\rho a^2 \qquad \text{.........(4).}$$

But from (1) we have

$$c^2 = \frac{4}{\pi} f\rho l^2 \qquad \text{.........(5),}$$

and from (4) and (5) we find

$$a > \frac{3}{\pi\sqrt{2}} l \qquad \text{.........(6).}$$

But

$$\tfrac{3}{4} b^2 l = a^3 > \frac{27}{\pi^3 2^{\frac{3}{2}}} l^3 \qquad \text{.........(7),}$$

whence

$$b > 0{\cdot}5l \qquad \text{.........(8).}$$

Hence the length and thickness of a section of the filament at rupture would be comparable, and there is no reason to suppose that sufficiently great changes took place in the ratio of its longitudinal and transverse dimensions to invalidate the supposition that the order of magnitude of the density remained unaltered during the adoption of the nearly spherical form.

Now substituting for a and l in (3) we find in turn

$$m > \frac{2^{\frac{1}{2}} 3^2}{\pi^2} \rho l^3 > \frac{3^2}{2^{\frac{5}{2}} \pi^{\frac{1}{2}}} \frac{c^3}{f^{\frac{3}{2}} \rho^{\frac{1}{2}}} \qquad \text{.........(9).}$$

This gives a lower limit to the mass of a planet that could be formed by gradual condensation of a gaseous filament. For nitrogen at 273° absolute the value of c is about 3×10^4 cm./sec. It is proportional to the square root of the absolute temperature and to the inverse square root of the molecular weight. Monatomic silicon would have the same molecular weight as diatomic nitrogen. The effective temperature and mean density of the sun were nearly as at present, say 5860° abs. and 1·4 gm./cm.^{3} Thus c for silicon at 5860° would be $3 \times 10^4 \left(\frac{5860}{273}\right)^{\frac{1}{2}}$ or $1{\cdot}4 \times 10^5$ cm./sec. The sun would be condensed towards the centre, and the ejected matter would be less dense than the average. We may take as working hypotheses

densities of 1·0 and 0·1 gm./cm.³ These give for the critical values, respectively,

$$m > 1{\cdot}5 \times 10^{26} \text{ or } > 4{\cdot}7 \times 10^{26}\,\text{gm.} \qquad \text{.........(10)},$$

$$a > \left(\frac{9c^2}{8\pi f\rho}\right)^{\frac{1}{2}} > 3{\cdot}3 \times 10^8 \text{ or } > 10{\cdot}4 \times 10^8\,\text{cm.} \qquad \text{.........(11)}.$$

For comparison we notice that the mass of Mercury is 2×10^{26} gm. and its radius $2{\cdot}2 \times 10^8$ cm. The theory thus appears to account for the formation of planets of the actual masses. The present radius must of course be less than a, which was the radius in the gaseous state. On the other hand a comes out much less than the radius of the sun, as it ought.

2·41. There is, however, a definite lower limit to the possible masses of planets produced by gradual condensation from matter ejected from a gaseous star. For, returning to equation 2·4 (2),

$$3c^2 < 2fm/a \qquad \text{.........(1)},$$

and supposing that when the planet ultimately solidified its radius was a' and its density ρ', we have

$$a > a' \qquad \text{.........(2)},$$

since the body must contract in passing from the gaseous to the solid state. Hence

$$3c^2 < 2fm/a' \qquad \text{.........(3)}.$$

If the density of the solid planet is taken as 3 gm./cm.³, m is nearly $12a'^3$, and therefore

$$24fa'^2 > 3c^2 \qquad \text{.........(4)},$$

which with the data so far adopted gives

$$a' > 2 \times 10^8\,\text{cm.} \qquad \text{.........(5)},$$

roughly. Thus no planet with a radius less than 2000 km. could have been formed by gradual condensation from the gaseous state.

If the density of the body when solid is 1 gm./cm.³, which is near the truth for many satellites, the critical radius is 3400 km., nearly that of the largest satellite of Jupiter.

This result is not dependent on any particular theory of the origin of the solar system, and the existence of many bodies within the system whose sizes fall below this limit establishes, as decisively as anything can be established in cosmogony, that these bodies were not formed by slow condensation from the gaseous state. They include all the asteroids, and all or nearly all the satellites. Thus at least two processes must have been operative in the formation of the planets and their satellites.

2·5. *The Condensation of the Planets*. The cooling of a condensation as massive as the great planets can now be described in some detail. Eddington has shown* that in a giant star a column of matter containing

* *M.N.R.A.S.* 77, 1917, 602.

one gram per square centimetre of cross section would absorb all but the fraction $e^{-5\cdot 4}$ of the radiation falling normally on its end. With the above data, radiation from the centre to the circumference would have to traverse $1\cdot 5 \times 10^6$ grams per square centimetre before reaching the outside, and therefore, unless the opacity dropped to a ten-millionth or less of its previous value as soon as it was ejected, it is justifiable to regard the radiation emitted by the filament and the primitive condensations as coming from a surface layer.

Imagine a sphere of radius a and effective temperature V; let its mean density be ρ and let us denote Stefan's constant by σ. Then the total rate of loss of energy by radiation is $4\pi\sigma a^2 V^4$, and the rate of loss per gram of matter is $3\sigma V^4/\rho a$. Taking

$$\begin{aligned} \sigma\,(1^\circ)^4 &= 5 \times 10^{-5} \text{ ergs/cm.}^2 \text{ sec.,} \\ V &= 5860^\circ, \\ \rho &= 1 \text{ gm./cm.}^3, \\ a &= 3\cdot 3 \times 10^8 \text{ cm.,} \end{aligned}$$

the rate of loss of energy is 536 ergs per gram per second, or $1\cdot 3 \times 10^{-5}$ calorie per gram per second. Thus the loss of heat would be at the rate of 420 calories per gram per year, which would liquefy the smaller planets in a few revolutions about the sun. Cooling would be delayed by internal atomic changes, energy set free by contraction, and solar radiation. All of these would be less effective in a gaseous planet than in the sun, and as they were in approximate equilibrium with radiation before the disruption, radiation would overwhelm them after it.

The mutual gravitation of the parts of a great planet would therefore hold it together, while radiation from the surface would gradually liquefy it. Since cooling would take place at the outside, drops would be formed there and would fall inwards under gravity. They would collect with the densest at the centre; solidification would in due course set in and proceed till complete.

2·51. Smaller masses of the same original density, or less dense masses of the same linear dimensions, would have a much more complicated history. In this case the attraction of the filament or of the condensation would be inadequate to retain the outer portions, which would therefore spread out with a velocity not exceeding $\sqrt{3}$ times the velocity of sound. Radiation would continue simultaneously. If A denote the area of a portion of the surface, the rate of loss of energy by radiation is $\sigma A V^4$. The temperature would fall to the boiling point by radiation still more quickly than in the case of a larger mass. Further radiation would produce liquefaction, and a volume of gas equal to $\sigma A V^4/\rho L$ would therefore liquefy in unit time, where L is the latent heat of evaporation. Meanwhile the rate of increase of volume through spreading would be $\sqrt{3}Ac$. If the former of these was

the greater, the reduction in volume through liquefaction would leave sufficient space to accommodate the expanding gases. The mass would therefore condense into liquid drops near the outside, and the still uncondensed gases would spread so as to fill only partially the space left by condensation. Thus expansion in volume would cease. The condition for this is that ρ shall be less than $\sigma V^4/\sqrt{3}Lc$, which with our previous data, taking L equal to 600 calories per gram, or 24×10^9 ergs per gram, is $1{\cdot}0 \times 10^{-5}$ gm./cm.^{3} If the matter was originally more dense than this, and gravity was insufficient to control its expansion, it would expand approximately adiabatically until it reached the critical density, when it would proceed to cool principally by radiation. If it cooled to the boiling point during the adiabatic expansion, much of it would have liquefied before it reached the critical density; if it had not cooled so far, it would liquefy at this density. This critical density is so much less than the probable density of the ejected matter that it is practically certain that adiabatic expansion would have produced liquefaction before the critical density was reached.

The drops formed would be at the same temperature as the gas that produced them. Now the velocity of agitation in a monatomic gas is $(\frac{9}{5})^{\frac{1}{2}}c$. The velocity of efflux is therefore 1·3 times the velocity of agitation. Thus the matter liquefying must be composed of molecules moving both inwards and outwards, for if they were all moving outwards their velocities could not differ enough among themselves to give the actual velocity of agitation. The effective front of the matter spreading out is composed chiefly of molecules moving forwards, and therefore the condensation must take place some distance behind the effective front; the latter goes straight on and is lost to the planet. Now the mass velocity of a drop after liquefaction would be the mean of the previous outward velocities of all its molecules, and since many of these would be negative, the drops would move outward with velocities much less than that of the effective front. If their velocity was still too great for them to be retained by the gravitation of the planet, they would pass off and be lost; but their inertia would delay the expansion of the gas following them, and thereby reduce the rate of expansion of the whole. It appears probable that the rate of expansion would ultimately be reduced to a very small fraction of the velocity of sound, and the condensation would become a system of liquid particles in an approximately stationary gaseous medium. The details of the process would be very difficult to follow out theoretically, but it appears possible that the velocity of expansion might be reduced to such an extent that gravity could control it. Thus adiabatic cooling leads to much the same result as radiative cooling would; the relation between them is that adiabatic cooling is the more important when the density is greater than the critical density, and that if the density ever falls to the critical value, or if it was initially below it, radiative cooling then takes the more prominent part.

2·52. In any mode of aggregation a planet would pass through a liquid stage. For the drops would be at the boiling point to start with, and would fall in towards the centre through a mass of gas at a temperature still far above their own, and the temperature of the gas would necessarily increase inwards on any theory of its thermodynamics. Thus their temperature would be raised both by conduction and by friction. When they reached the centre they would come to rest, being further heated by their mutual impacts, and would form a liquid core. Thenceforth the core would continue to be heated by contact with the surrounding gas. This would partially condense on the core, which would itself be thereby heated to a temperature not greater than the boiling point at the actual pressure. The outer gases might either condense, the drops falling in to augment the nucleus, or else pass off into space; but it is certain that any matter aggregated into the nucleus would be liquid.

The last point is of capital importance in the development of geophysical theory, and therefore merits further consideration, which may conveniently be given here. It has been asserted by Chamberlin and others* that the drops would cool not only to the liquid but to the solid state, and that the nucleus of a small planet would accordingly have been initially and permanently solid. The objection is that a liquid particle falling through gas would necessarily be heated; if therefore it was to cool further it would have to emerge from the region occupied by gas. Thus it would have to continue to move outwards with a velocity greater than the velocity of expansion of the gas, which by hypothesis is itself too great to be controlled by gravity. Thus such a particle would be lost to the condensation. If on the other hand the planet was of such mass that it could control particles moving at the surface with the velocity of efflux of a gas into a vacuum, it is certain that liquid particles would commence to fall inwards as soon as they were formed. Thus in the case of a large planet the formation of a liquid core is certain; in the case of a lighter planet the alternatives are either a planet formerly fluid or no planet at all.

It has also been thought by Chamberlin and others that adiabatic cooling would proceed after condensation and continue till the mass was solid and cold. This is impossible, by what has been said above, since the surroundings would always be as hot as the core. But even if free cooling of the drops was possible, they would still not be cooled to the solid state by adiabatic cooling. For adiabatic cooling below the boiling point can be caused only by evaporation, and therefore could not lower the temperature below a point where the vapour pressure is insignificant. Thus the temperature could never be reduced in this way by more than 200° at most below the boiling point. But the difference between the boiling points and melting points of silicon and heavy metals is at least several hundred degrees. Hence adiabatic cooling could reduce the

* See Appendix A.

substance of the planets at most to the liquid state and not to the solid state. If evaporation was appreciable in substances with high melting points when fused, the casting of iron and the melting of glass should produce a great amount of distillation of iron and glass on to the factory walls, which does not take place.

2·53. Returning now to the question of the disruptive action of the sun we see that it could not impede condensation. For it was virtually taken into account by Jeans in finding the condition that the filament might break up into detached masses, and the length of the segments obtained by him is practically determined by the condition that the matter may be far enough from the sun not to be broken up further. As it receded from the sun, it would become more nearly spherical and more dense, and on both grounds the sun's disruptive influence would diminish, apart from the direct effect of distance.

2·54. In this way a number of liquid planets of varying sizes would be formed. They would move in orbits about the sun, all in one direction and approximately in one plane. Their outward motion would not in general be annulled during the passage of the star, so that after the encounter they would have considerable velocities away from the sun; in other words, their orbits would be highly eccentric.

2·6. *Buffon's Theory.* At this stage we may quote the passage from Laplace's *Système du monde* that was omitted on p. 6. It is as follows:

Buffon is the only person I know who, since the true system of the world was discovered, has attempted to find out the origin of the planets and satellites. He supposes that a comet, falling upon the sun, drove from it a torrent of matter, which united far away into several globes of various sizes and at various distances from that body. These globes are the planets and satellites, which, by cooling, have become opaque and solid.

This hypothesis satisfies the first of the five conditions already mentioned*; for it is clear that all the bodies thus formed must move nearly in the plane that included the centre of the sun and the path of the torrent of matter that formed them. The four other phenomena appear to me inexplicable by this means. In truth, the absolute movement of the molecules of a planet must be in the direction of the motion of its centre of gravity, but it does not follow that the rotation of the planet will be in the same sense. Thus, the earth could turn from east to west, and yet the absolute movement of each of its molecules could be from west to east. This applies also to the motion of revolution of the satellites, whose direction, on this hypothesis, is not necessarily the same as that of the motion of their primaries.

The smallness of the eccentricities of the planetary orbits is not only very difficult to explain on this hypothesis, but actually contrary to it. We know from the theory of central forces that if a body, moving in a closed orbit about the sun, touches the surface of that luminary, it will return there in every revolution. Hence it follows that if the planets had been primitively detached

* See 1·1.

from the sun, they would touch it at each revolution, and their orbits, instead of being circular, would be very eccentric. It is true that a torrent of matter expelled from the sun cannot be exactly compared to a globe grazing the surface: the pressure and the gravitational attraction between the parts of the torrent may change the directions of their movements and make their perihelia recede from the sun. But their orbits must remain permanently very eccentric; or, at least, they could have had small eccentricities only by the most extraordinary chance. Finally, one sees no reason, on Buffon's hypothesis, why the orbits of the comets already observed, about ninety in number, are all very much elongated; this hypothesis is therefore far from satisfying our conditions.

If in Buffon's hypothesis we replace 'comet' by 'star more massive than the sun' and 'falling upon the sun' by 'approaching very close to the sun,' we have a hypothesis with a close resemblance to the one that has been elaborated in this chapter. We must, therefore, be prepared to meet Laplace's objections. We note that his second point is certainly false, eight retrograde satellites being now known; and the third point is untrue of Uranus and probably of Neptune. The fifth remains a difficulty, comets not having yet been satisfactorily included in any cosmogonical hypothesis (Laplace's own being no exception). The fourth point has now been met*, and it appears that the present small eccentricities of the planetary orbits are perfectly consistent with the tidal theory.

2·7. It will be noticed that in the primitive sun the lightest materials would be in the outer layers; they would therefore be the first ejected, and would proceed to the greatest distances from the sun. Thus the outermost planets would be expected to be composed of the lightest materials and to have the lowest densities, as they have. If any matter was expelled from the solar system altogether, as is quite possible, it would be the lightest of all. The effect would, however, have been intensified by the fact that the lighter planets would have lost much more of their lighter materials, and would thus have come to contain a greater proportion of heavy constituents than they originally possessed.

It has been seen that much expelled matter must have fallen back into the sun, taking with it the angular momentum that it acquired during its journey. Hence the sun acquired a rotation, the plane of its equator being near the planes of the motions of the planets, as it actually is.

2·8. *The Origin of Satellites.* It appears unlikely that all the bodies in the solar system were produced in the disruption already discussed. The diameter and density of the filament would certainly vary from point to point, but it is incredible that they could vary in such an irregular way as to account for the occurrence of bodies with such widely different masses as Saturn and its satellites in close proximity and between two other bodies both comparable in mass with Saturn. This is to take only one example of the difficulties presented by the wide differences in mass

* See Chapter IV.

between the outer planets and their satellites. A natural suggestion is that the satellites were formed from the planets. They could not have been formed by gradual condensation of the planets, for every argument used against the Laplacian theory of the origin of the planets and against its modifications is equally applicable to the corresponding theories of the origin of the satellites from the planets. The tidal theory, on the other hand, is applicable to this problem also. Each nucleus would pass near the sun at its first perihelion after it was formed. How near to the sun it would pass presents a perhaps quite intractable question in the Problem of Three Bodies; but it is at least plausible that the perihelion distance would be less than the distance of its centre from the sun when it first took a spherical form. The latter distance is specified by the condition that the body could just hold itself together in spite of the disruptive tidal action of the sun. If it had not condensed appreciably during its first revolution it would therefore be broken up by solar tidal action at its first perihelion. If it approached more closely it might be broken up even in spite of the condensation. A filament would then be produced by the planet, and would go through a process of development similar to that of the filament ejected by the sun. This is the mode of origin of the satellites suggested by Jeans. It is not, however, a complete account. Just as the planets necessarily move in one direction around the sun, any satellite would necessarily move in that direction about its primary if it had been produced in this way and always remained with its primary. But just as in the first great disruption much matter might have acquired a great enough velocity to expel it from the sun's influence altogether, so in these minor disruptions some satellites formed might have been permanently lost to their primaries and proceeded to describe independent orbits about the sun. It is possible that in their later development they would be captured* by their own parents or by other planets, or they might remain permanently independent.

The smaller planets may have become largely liquid before passing perihelion for the first time. If so, Jeans's theory of the tidal disruption of incompressible masses would apply to them†. Equation (102) of his paper implies that a fluid planet could exist close to the sun's surface if its density exceeded 10·6 times the sun's mean density, or 15 gm./cm.^{3} A sufficiently close approach to the sun would break up any planet even in its present state, but the approach would have to be very close for the dense terrestrial planets. Tidal disruption of a liquid body would give a satellite or satellites comparable in size with the primary. Such a theory therefore offers hope of accounting for the moon, but not for the satellites of Mars.

This argument is not applicable to the great planets, for they would

* See later, Chapter IV.

† *Memoirs R.A.S.* 62, 1917, 35.

be still distended at their first perihelion passage, and would not reach a density that would forbid disruption until after a few revolutions.

It has been pointed out that the falling back into the sun of temporarily expelled matter would account for the sun's rotation; similarly the rotation of the planets may be explained. It is suggestive that the only planets with swift direct rotations are those whose direct-moving satellites indicate that they have been broken up by tidal action. The angular momenta of revolution of the satellites of the great planets are, however, much less than those of the rotations of the planets, so that this hypothesis cannot be accepted without a great deal of examination.

With regard to the method of condensation of the bodies produced, it is probable that all the planets solidified by the process of 2·5 and all the satellites by that of 2·51. But the terrestrial planets have densities several times those of the great planets and their satellites, suggesting that they lost much of their lighter constituents during condensation.

2·9. *Summary.* A theory of the origin of the solar system, partly based on that of Jeans, has been developed. The sun, while in the giant stage, is supposed to have been broken up by the tidal action of a passing star several times more massive than itself. It is unlikely that the sun's size, density, and temperature at the time of the encounter differed to any notable extent from their present values. The encounter did not approximate either to Jeans's 'slow' or to his 'transitory' type. A slow encounter is dynamically impossible, while the matter ejected in a true transitory encounter would all fall back into the sun and be reabsorbed.

Accordingly the encounter must have been of intermediate type. The ejected matter, as it emerged from the sun, collected into nuclei. These continued to move outwards, but were deflected sideways by the star, and thus proceeded to move around the sun in one direction and nearly in one plane. The mass of the smallest nucleus able to retain its chief present constituents must have been comparable with that of Mercury. The great outer planets must therefore have retained all their constituents, while the smaller planets have probably lost a large part of their lighter ones; thus the fact that the great planets have low densities may be explicable.

The condensation of the great planets was a straightforward process, each passing steadily through the gaseous state to the liquid state through loss of heat by radiation from the outside. The smaller ones and the satellites had a more complicated history. They would form drops at the outside, and these would fall in towards the centre, forming liquid cores; but at the same time a large fraction of their mass would probably be lost.

When the terrestrial planets had reached the liquid state they might continue to cool adiabatically by evaporation from the surface; but evaporation alone would not have sufficed to bring them to the solid state,

for it would have become inappreciable before they had cooled so far. Thus all the planets must have gone through a liquid stage, passing gradually into the solid state by cooling from the surface.

Most of the satellites were probably formed by the tidal disruption of their primaries by the sun when they passed perihelion for the first time. This hypothesis accounts readily for the general resemblance of the subsystems of the great planets to the solar system as a whole, and for the dissimilarity of the subsystems of the terrestrial planets. It suggests also that a number of satellites were detached completely from their original parents and became for a time independent planets. An alternative explanation is available for the origin of the moon, as will be seen in Chapter III. The possible subsequent history of the lost satellites will be discussed in 4·4.

CHAPTER III

The Origin of the Moon

> "There is a tide in the affairs of men
> Which taken at the flood leads on to fortune."
>
> SHAKESPEARE, *Julius Caesar*, IV, 3.

3·1. Our moon has probably had a very different origin and history from any other satellite. Although the earth is the second smallest planet to possess a satellite at all, the moon is the fifth most massive satellite in the whole system. Its mass is $\frac{1}{82}$ of that of the earth, while the best estimates available of the masses of other satellites indicate that the only heavier ones are J I, J III, J IV, and Titan, whose masses are respectively $\frac{1}{70}$, $\frac{1}{40}$, $\frac{1}{70}$, and $\frac{1}{60}$ of that of the earth. The ratios of the masses of these large satellites to those of their primaries are $\frac{1}{22200}$, $\frac{1}{12500}$, $\frac{1}{22200}$, and $\frac{1}{4700}$*. We have seen that if the earth had become liquid and later been broken up by tides raised by the sun, a relatively large satellite would be expected. In the first edition of this work this possibility was excluded, because it was assumed throughout that the sun at the time was very distended. The earth when liquid cannot have been much more bulky than at present, and we have seen that the sun could not have broken it up unless the density of the sun exceeded $\frac{1}{10}$ of that of the earth. This was inconsistent with the hypothesis of any notable distension of the sun. But now that we are regarding the primitive sun as having had nearly its present density, tidal disruption of the liquid earth becomes possible. It is still, however, hardly probable, because the approach required must be within about 1·4 solar radii of the centre of the sun.

Of the other alternatives, it is perhaps just possible that the moon was originally an independent planet, though it is much less massive than any existing planet. It cannot have been formed from any planet other than the earth; not from another terrestrial planet, because this would intensify the difficulty with regard to its mass; not from an outer planet, because a lost satellite would be expected to be of lighter materials than the retained satellites, and actually the moon is far the densest satellite in the system.

3·2. There remains the possibility that the moon was actually formed from the earth by the action of the solar tides, but with the aid of another circumstance. Such a theory was suggested by Sir G. H. Darwin†. Consider for a moment a man in a swing. If he is pulled back for a moment

* Cf. R. A. Sampson, *Memoirs R.A.S.* **63**, 1921, 178.

† *Phil. Trans.* **170**, 1879, Part II, 537.

and released, he performs a few oscillations and gradually comes to rest. But if we give him a push every time he is nearest to us, the extent of the oscillations becomes greater every time, and if friction were absent, it would be possible to increase the amplitude indefinitely. This increase in the extent of a vibration when the external force has a period equal to the natural period is known as 'resonance.' If the earth, when it was wholly or partially liquid, received a small distortion so that its equator became an ellipse, it would oscillate backwards and forwards about the symmetrical form until friction brought it to rest. The period of this motion would be about two hours. But we can show that the angular momentum of the earth-moon system is such that, if the earth and moon ever formed a single body, this must have rotated in about four hours. Now the solar semidiurnal tide is just such an oscillation as we have been considering, and the period of the disturbance in height at any station is half the period of rotation, in this case two hours. Hence the amplitude of the oscillation would grow to be very great. There appears to be no limit to the amplitude of the tide that could be produced in this way. For suppose the oscillation to have continued long enough for its amplitude to have become steady, and consider the elevation of the surface at the point vertically below the sun. If the free period was shorter than the period of the tide, the tidal elevation would be positive, in accordance with the ordinary theory of forced oscillations; but if the period of the tide was the shorter, the elevation would be negative. Hence if the earth was condensing slowly, both periods changing slowly, and the one that was formerly the shorter became the longer, the elevation would change its sign. It could do this only by passing through zero or infinity. The former alternative implies that when perfect resonance is attained there would be no tide, which is unplausible; the latter implies that if only the conditions for resonance persisted for a long enough time there is no limit to the extent of the tide that could be produced. Thus the earth would be gradually stretched out in the direction of the sun. When the disturbance became great enough, the mass would break into two parts in much the same way as Jeans showed to be possible for a homogeneous liquid mass undergoing a slow tidal encounter without resonance.

Love* and Bryan† determined the angular velocity that the earth would have had to have in order to produce a satellite in this way, supposing the earth to have been homogeneous, but Moulton‡ showed that the angular momentum in the system is too small to give the velocity of rotation required. It was found by the present writer, however, that when we take into account the fact that the earth is not homogeneous, the conditions become much more favourable to the theory. The oscillations of a heterogeneous mass in three dimensions have so far presented an

* *Phil. Mag.* (5), **27**, 1889, 254–264. † *Phil. Trans.* **153** A, 1889, 187–219.

‡ T. C. Chamberlin and others, *The tidal and other problems*, Carnegie Institute.

intractable problem, but the corresponding two-dimensional problem has been solved, and it is found that homogeneity is the least favourable case to the theory that is possible. The modification produced in the numerical results by allowing as much heterogeneity in the two-dimensional problem as exists in the actual earth is found to be as great as is required to remove the discrepancy found by Moulton.

3·21. *The Free Vibrations of a Heterogeneous Liquid Cylinder.* In the equilibrium state the liquid is supposed to form two layers, an inner circular cylinder of density $\rho\,(1 + \eta)$ and radius $a_1 = \alpha a$, surrounded by an outer layer of density ρ and radius a. Let the axis of z be along the centre of the cylinder, and suppose a constraint to prevent motion along the cylinder. Let the angular velocity be ω. The equations of motion, referred to rectangular axes rotating with this speed, are

$$\left.\begin{aligned}\dot{u} - 2\omega v - \omega^2 x &= \frac{\partial}{\partial x}\left(U - \frac{p}{\rho}\right)\\ \dot{v} + 2\omega u - \omega^2 y &= \frac{\partial}{\partial y}\left(U - \frac{p}{\rho}\right)\end{aligned}\right\} \qquad \text{.........(1)},$$

where u and v are the rates of change of the coordinates of a given particle, p is the pressure, and U the gravitation potential; u and v are supposed small. The equation of continuity is

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0 \qquad \text{.........(2)}.$$

By cross-differentiation we at once find

$$\frac{\partial}{\partial t}\left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right) = 0,$$

so that in all periodic motions

$$\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} = 0 \qquad \text{.........(3)}.$$

Hence there are a velocity potential Φ and a stream function Ψ giving the motion relative to the rotating axes. Then

$$\nabla^2\Phi = 0 \text{ and } \nabla^2\Psi = 0 \qquad \text{.........(4)}.$$

This will hold for both layers.

Let the equations of the bounding layers in cylindrical polar coordinates in the disturbed position be respectively

$$\left.\begin{aligned}r &= a + q\cos(n\phi + \beta) = a + qS, \quad \text{say}\\ r &= a_1 + q_1\cos(n\phi + \gamma) = a_1 + q_1 S_1, \quad \text{say}\end{aligned}\right\} \qquad \text{.........(5)},$$

where β and γ may be functions of the time and q and q_1 are small.

Then the gravitation potentials in the outer and inner layers are

$$U_1 = -\pi f\rho (r^2 - a^2) - 2\pi f\rho\eta a_1^2 \log\frac{r}{a_1} + \frac{2\pi f\rho}{n}\left\{aqS\left(\frac{r}{a}\right)^n + \eta a_1 q_1 S_1 \left(\frac{a_1}{r}\right)^n\right\} \quad \ldots(6),$$

$$U_2 = -\pi f\rho (r^2 - a^2) - \pi f\rho\eta (r^2 - a_1^2) + \frac{2\pi f\rho}{n}\left\{aqS\left(\frac{r}{a}\right)^n + \eta a_1 q_1 S_1 \left(\frac{r}{a_1}\right)^n\right\} \quad \ldots(7).$$

When $r = a$, let the part of $-\frac{\partial U}{\partial r}$ that is independent of ϕ be g; and when $r = a_1$, let the part independent of ϕ be g_1.

We have further

$$U - \frac{p}{\rho} = \int (\dot{u} - 2\omega v - \omega^2 x)\, dx + (\dot{v} + 2\omega u - \omega^2 y)\, dy$$

$$= \frac{\partial \Phi}{\partial t} - \tfrac{1}{2}\omega^2 r^2 + 2\omega\Psi \quad \ldots\ldots\ldots(8).$$

Assume next that in the inner and outer layers the velocity potentials have the forms

$$\Phi_1 = \left(\frac{r}{a}\right)^n K_1 + \left(\frac{a}{r}\right)^n K_2 \quad \ldots\ldots\ldots(9),$$

$$\Phi_2 = \left(\frac{r}{a_1}\right)^n K_3 \quad \ldots\ldots\ldots(10),$$

where K_1, K_2, and K_3 are harmonic functions of $n\phi$. Then the stream functions are obtained by writing

$$n\phi - \frac{\pi}{2} \text{ for } n\phi \text{ in } K_1 \text{ and } K_3, \text{ and } n\phi + \frac{\pi}{2} \text{ for } n\phi \text{ in } K_2 \ldots\ldots(11).$$

Then the boundary conditions give

$$n(K_1 - K_2) = a\frac{d}{dt}(qS) \quad \ldots\ldots\ldots(12),$$

$$n(K_1\alpha^n - K_2\alpha^{-n}) = a\alpha\frac{d}{dt}(q_1 S_1) = nK_3 \quad \ldots\ldots\ldots(13).$$

Hence

$$K_3 = \frac{a\alpha}{n}\frac{d}{dt}(q_1 S_1) \quad \ldots\ldots\ldots(14),$$

$$K_1 = \frac{d}{dt}\frac{a(qS - q_1 S_1 \alpha^{n+1})}{n(1 - \alpha^{2n})};\ K_2 = \frac{d}{dt}\frac{a(qS\alpha^{2n} - q_1 S_1 \alpha^{n+1})}{n(1 - \alpha^{2n})} \quad \ldots(15).$$

At the outer surface $\Psi = \frac{d}{dt}\frac{aq}{n} S\left(n\phi - \frac{\pi}{2}\right)$;

at the inner surface $\Psi = \frac{d}{dt}\frac{a\alpha}{n} q_1 S_1 \left(n\phi - \frac{\pi}{2}\right)$.

At the outer surface the pressure is constant. Therefore

$$U - \frac{\partial \Phi}{\partial t} + \omega^2 aqS - 2\omega\Psi = \text{constant},$$

or $-gqS + \frac{2\pi f\rho a}{n}(qS + \eta\alpha^{n+1} q_1 S_1) + \omega^2 aqS$

$$- \frac{d^2}{dt^2}\frac{a\{qS(1 + \alpha^{2n}) - 2\alpha^{n+1} q_1 S_1\}}{n(1 - \alpha^{2n})} - \frac{2\omega}{n}\frac{d}{dt} aqS\left(n\phi - \frac{\pi}{2}\right) = 0 \quad \ldots(16).$$

At the inner surface the pressure is continuous. Hence

$$\eta\left[-g_1q_1S_1+\frac{2\pi f\rho a}{n}(qSa^n+\eta q_1S_1a)+\omega^2 aaq_1S_1-2\omega\frac{aa}{n}\frac{d}{dt}q_1S_1\left(n\phi-\frac{\pi}{2}\right)\right]$$
$$-(1+\eta)\frac{aa}{n}\frac{d^2}{dt^2}(q_1S_1)+\frac{d^2}{dt^2}\frac{a\{2qSa^n-q_1S_1a(1+a^{2n})\}}{n(1-a^{2n})}=0\quad\text{......(17).}$$

So far nothing has been assumed about the form of the wave, save that S and S_1 are circular functions of the nth order. Suppose now that the wave travels with its form unchanged and that the speed of the vibration at any point is p. Then q and q_1 are constants, and S and S_1 are of the form $\cos(n\phi-pt+\beta)$.

$$\text{Then }\frac{d}{dt}S\left(n\phi-\frac{\pi}{2}\right)=-pS,\text{ and }\frac{d}{dt}S_1\left(n\phi-\frac{\pi}{2}\right)=-pS_1\quad\text{......(18).}$$

$$\text{Thus}\quad qS\left[-\frac{ng}{a}+2\pi f\rho+n\omega^2+p^2\frac{1+a^{2n}}{1-a^{2n}}+2\omega p\right]$$
$$+q_1S_1\left[2\pi f\rho\eta a^{n+1}-\frac{2p^2a^{n+1}}{1-a^{2n}}\right]=0\quad\text{.........(19),}$$

$$qS\left[2\pi f\rho\eta a^{n-1}-\frac{2p^2a^{n-1}}{1-a^{2n}}\right]$$
$$+q_1S_1\left[-\eta\frac{ng_1}{aa}+2\pi f\rho\eta^2+n\eta\omega^2+(1+\eta)p^2+p^2\frac{1+a^{2n}}{1-a^{2n}}+2\eta\omega p\right]=0$$
$$\text{.........(19 }a\text{).}$$

The period equation is therefore

$$\left(-\frac{ng}{a}+2\pi f\rho+n\omega^2+2\omega p+p^2\frac{1+a^{2n}}{1-a^{2n}}\right)$$
$$\left(-\frac{\eta ng_1}{aa}+2\pi f\rho\eta^2+\eta n\omega^2+2\eta\omega p+\eta p^2+\frac{2p^2}{1-a^{2n}}\right)-\left(\pi f\rho\eta-\frac{p^2}{1-a^{2n}}\right)^2 4a^{2n}=0$$
$$\text{.........(20).}$$

In the homogeneous case $\eta=0$, and this equation reduces to

$$p^2+2\omega p+n\omega^2+2\pi f\rho-ng/a=0\ \text{ or }\ p^2=0\quad\text{.........(21).}$$

The zero roots give merely a displacement of the inner boundary without the outer being affected, which is obviously possible when the densities are equal.

In the other type of oscillation the sum of the two possible speeds is -2ω, as was proved by Bryan for the corresponding three-dimensional problem. Further, in the case of the form of bifurcation ($n=2$, $p=0$) $\omega^2=\pi f\rho$, as was found by Jeans*.

We wish to know whether heterogeneity causes resonance or instability to occur for a smaller value of the angular velocity when the mean density is the same. Now the mean density is $\rho(1+\eta a^2)$, and hence the angular velocity that would cause both these conditions in a homogeneous cylinder with the same mean density is given by

$$\omega_1{}^2=\pi f\rho(1+\eta a^2)\quad\text{.........(22).}$$

* 'The Equilibrium of Rotating Liquid Cylinders,' *Phil. Trans.* **200 A, 1903, 81.**

Resonance will occur in the heterogeneous case if one value of p is -2ω. In this case the angular velocity ω_2 is given by

$$\frac{8\omega^2}{1-\alpha^{2n}}\left\{-\frac{ng}{a}-\frac{\eta n g_1}{a}\alpha^{2n-1}+2\pi f\rho\,(1+2\eta\alpha^{2n}+2\eta^2\alpha^{2n})+n\omega^2\,(1+\eta\alpha^{2n})\right\}$$
$$+\eta\left(n\omega^2+2\pi f\rho-\frac{ng}{a}\right)\left(n\omega^2+2\pi f\rho\eta-\frac{ng_1}{a\alpha}\right)-4\pi^2 f^2\rho^2\eta^2\alpha^{2n}=0 \quad \ldots(23).$$

Instability will commence with an angular velocity ω_3 given by putting $p=0$. Then

$$\left(n\omega^2+2\pi f\rho-\frac{ng}{a}\right)\left(n\omega^2+2\pi f\rho\eta-\frac{ng_1}{a\alpha}\right)-4\pi^2 f^2\rho^2\eta\alpha^{2n}=0 \quad \ldots\ldots(24).$$

As we are restricted to a two-dimensional problem and only require approximate results to indicate the direction of the effect of heterogeneity, accurate correspondence with the actual case is unnecessary. We shall assume

$$\rho=3{\cdot}2,\quad \eta=1{\cdot}50,\quad \alpha=0{\cdot}66.$$

Then $$\rho\,(1+\eta)=8{\cdot}0;\ g/a=3{\cdot}30\pi f\rho;\ g_1/a\alpha=5{\cdot}00\pi f\rho \quad \ldots\ldots\ldots(25).$$

This makes the two densities, and the ratio of the amounts of matter of the two densities, about the same as in Wiechert's hypothesis of the structure of the earth. Then

$$\omega_1{}^2/\pi f\rho=1{\cdot}65 \quad \ldots\ldots\ldots(26).$$

Putting $n=2$ we find

$$\omega_2{}^2/\pi f\rho=0{\cdot}71 \text{ or } 2{\cdot}10 \quad \ldots\ldots\ldots(27),$$
$$\omega_3{}^2/\pi f\rho=2{\cdot}03 \text{ or } 3{\cdot}74 \quad \ldots\ldots\ldots(28).$$

If a widely diffused cylindrical mass with the density distributed in this way condenses so as to keep α constant, ρ and ω increase like a^{-2}. Thus $\omega^2/\pi f\rho$ increases like a^{-2}, and resonance will occur when it reaches the value 0·71, which is only 0·43 of the value needed to produce resonance and instability in a homogeneous mass with the same mean density. Thus heterogeneity encourages resonance; the fact that the smaller value of $\omega_3{}^2$ is greater than $\omega_1{}^2$ indicates that it discourages instability.

In general, put $p=k\omega$ and $\omega^2/\pi f\rho=\lambda$. Then the period equation for $n=2$ is

$$\left\{-(1+2\eta\alpha^2)+\lambda\left(1+k+\tfrac{1}{2}k^2\frac{1+\alpha^4}{1-\alpha^4}\right)\right\}$$
$$\left\{-(2+\eta)+\lambda\left(1+k+\tfrac{1}{2}k^2+\frac{k^2}{\eta\,(1-\alpha^4)}\right)\right\}\eta-\left\{\eta-\frac{k^2\lambda}{1-\alpha^4}\right\}^2\alpha^4=0 \ldots(29).$$

This may be regarded as a quadratic in λ when k is known. The solution is given in the table. In the last two columns λ_0 is the value of $\omega^2/\pi f\rho$ for a homogeneous cylinder of density $\rho\,(1+\eta\alpha^2)$, so that a direct comparison is obtained between heterogeneous and homogeneous cylinders of the same mean density.

k	λ		$\lambda^{\frac{1}{2}}$		$k\lambda^{\frac{1}{2}}$		$\lambda_0^{\frac{1}{2}}$	$k\lambda_0^{\frac{1}{2}}$
∞	0	0	0	0	1·53	2·06	0·00	1·82
3	0·200	0·254	0·447	0·504	1·34	1·51	0·44	1·33
2	0·367	0·423	0·606	0·650	1·21	1·30	0·57	1·15
1	0·83	1·05	0·91	1·03	0·91	1·03	0·81	0·81
0·5	1·30	1·95	1·14	1·40	0·57	0·70	1·01	0·50
0	2·03	3·74	1·42	1·93	0·00	0·00	1·28	0·00
−0·5	2·83	5·43	1·68	2·33	−0·84	−1·12	1·65	−0·83
−1·0	2·34	4·23	1·53	2·06	−1·53	−2·06	1·82	−1·82
−2·0	0·71	2·10	0·84	1·45	−1·68	−2·90	1·28	−2·57
$-\infty$	0	0	0	0	−1·53	−2·06	0·00	−1·82

In the diagram the full curves show the variation of $k\lambda^{\frac{1}{2}}$ with $\lambda^{\frac{1}{2}}$, or for masses with the same densities, of p with ω. The two branches of the curve approach very closely near $k = 2$, but do not intersect. The dotted curve is obtained by plotting $k\lambda_0^{\frac{1}{2}}$ against $\lambda_0^{\frac{1}{2}}$, so that it shows the speeds on the same scale for masses of the same mean density. The condition for resonance is given by the intersection of the curves with the line $p = -2\omega$ or $k = -2$, and it is at once seen that heterogeneity causes this to be satisfied for a much smaller value of the rotation than was otherwise needed. Further, if a be made to approach to unity, two of the values of k can be made to approach zero as closely as we like, while λ retains any finite value. Hence by making the depth of the outer layer small enough we can make resonance occur for as small a value of the rotation as we please. It appears physically probable that the same will be true in the three-dimensional problem.

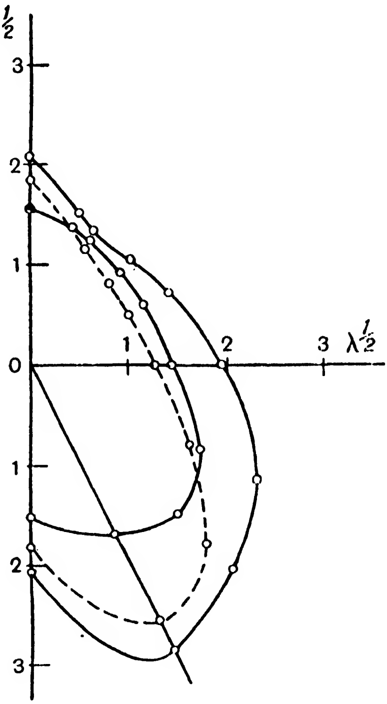


Fig. 3

3·3. *The Original Period of Rotation.* If C is the earth's greatest moment of inertia, M its mass, and a its radius,

$$C = 0{\cdot}334Ma^2 \qquad \text{.........(1).}$$

Let ω and n be the respective angular velocities of the earth's rotation and the moon's revolution at present; let c be the moon's distance and m its mass. Then the angular momentum of the system is

$$C\omega + \frac{Mm}{M+m}c^2n = 5{\cdot}78C\omega \qquad \text{.........(2).}$$

Now $\omega = 7{\cdot}29 \times 10^{-5}/1$ sec. The moment of inertia of the combined body before separation would be approximately $C\,(M + m)^{\frac{5}{3}}\, m^{-\frac{5}{3}}$, apart from the increase due to flattening. Hence the angular velocity when the two formed one body was

$$5{\cdot}78\omega \left(\frac{M}{M+m}\right)^{\frac{5}{3}} = 4{\cdot}14 \times 10^{-4}/1 \text{ sec.},$$

provided that the density distribution was similar, and that the angular momentum and moment of inertia were not different. The only cause of change in angular momentum would be solar tidal friction, which probably would not amount to more than about 5 per cent. of the whole. Thus

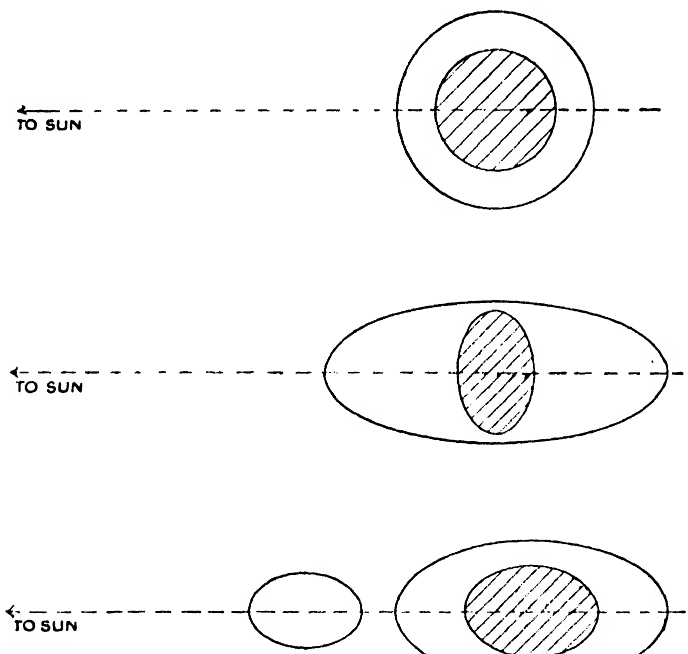


Fig. 4. Changes in form of earth during resonance.
(From *Evolution: A Collective Work*. Blackie, 1925.)

finally the angular momentum would be enough to make the whole rotate with an angular velocity of about $4{\cdot}3 \times 10^{-4}/1$ sec., apart from variations in the moment of inertia due to flattening and condensation.

Now a homogeneous ellipsoid would give similarly a maximum angular velocity of $3{\cdot}6 \times 10^{-4}/1$ sec., and instability in the symmetrical form and resonance would not occur until the rotation speed was $7{\cdot}2 \times 10^{-4}/1$ sec. Thus the homogeneous mass could never attain conditions suitable for resonance. The effect of heterogeneity on these conditions seems to be to increase the actual speed, by what has been said above, and to diminish the speed needed for resonance, so that the circumstances are much more favourable.

The problem of finding out the actual amount of the effect of heterogeneity on the period of rotation necessary to cause resonance is likely to be exceedingly difficult, as the bounding surfaces are not ellipsoids. An estimate can, however, be made by analogy with the cylindrical case, which has been accurately solved. In the two-dimensional case considered the angular velocity found necessary for resonance is only 0·65 of that needed when the mass is homogeneous. If the same ratio held in the three-dimensional case, the angular velocity needed would be $4{\cdot}7 \times 10^{-4}/1$ sec., while the available angular velocity is $4{\cdot}3 \times 10^{-4}/1$ sec. There are, however, two causes that will affect the former amount. Compressibility by introducing a further degree of freedom may be expected to reduce the free speed for any given angular velocity, and hence to reduce the angular velocity needed to make the ratio k equal to -2. Further, by reducing the relative thickness of the outer layer this free speed may be indefinitely reduced, and therefore it is evident that there can be a heterogeneous distribution of density, not very different from the actual distribution, such that with the same mean density the available angular momentum would lead to resonance. With still smaller relative thicknesses resonance would still be possible, but would occur at an earlier stage of condensation. It is therefore highly probable that conditions suitable for resonance occurred.

3·4. *The Vibrations of a System with one degree of Freedom when the Free and Forced Periods are nearly equal and slowly varying.* Let the equation of motion of the system be

$$\ddot{x} + \mu\dot{x} + a^2x = Ee^{\iota pt} \qquad \text{.........(1)},$$

where μ is small, and a, p, and μ are slowly varying.

Put $\int a dt = w$, and let accents denote differentiation with regard to w.

Then
$$x'' + \nu x' + x = Ea^{-2} \exp \iota p\int a^{-1}dw \qquad \text{.........(2)},$$

where
$$\nu = \frac{1}{a}\frac{da}{dw} + \frac{\mu}{a} \qquad \text{.........(3)}.$$

Now let one of the complementary functions of this be $\exp \theta$. Then it has been already shown* that the value of θ is practically $(\iota - \frac{1}{2}\nu)\, w$ for a considerable range in w.

Put $x = y \exp \theta$. Then

$$y'' + y'\,(2\theta' + \nu) = \mathrm{E}a^{-2} \exp \left(\iota p \int \frac{dw}{a} - \theta\right) \qquad \text{.........(4)}.$$

If p and a are nearly equal, we can write $p\displaystyle\int \frac{dw}{a} = (1 + \beta w)\, w$, where β is small. Then the expression on the right of (4) is $Ea^{-2} \exp\,(\frac{1}{2}\nu w + \iota\beta w^2)$,

* *Memoirs of R.A.S.* **60**, 1915, 211–213.

which varies slowly. Hence y'' can be neglected on the left, and the particular integral is approximately given by

$$y = \frac{E\alpha^{-2}}{2\iota}\int \exp\left(\tfrac{1}{2}\nu w\right) dw = \frac{E\alpha^{-2}}{\iota\nu}\exp\left(\tfrac{1}{2}\,\nu w\right)$$

near the instant when the periods coincide.

Hence

$$x = \frac{E\alpha^{-2}}{\iota\nu}\exp(\iota w) + A\exp\left(\iota - \tfrac{1}{2}\nu\right)w + B\exp - \left(\iota + \tfrac{1}{2}\nu\right)w \ldots\ldots(5),$$

where A and B are arbitrary constants. The amplitude of the forced vibration is therefore magnified in the ratio $\frac{1}{\nu}$, or $\alpha\Big/\left(\frac{d\alpha}{dw} + \mu\right)$ when the periods coincide.

The equilibrium amplitude of the solar semidiurnal tide is about 20 cm., if the mutual attraction of the matter of the earth is included. For resonance to lead to rupture this must be magnified to something of the order of 3000 km. Thus the magnification needed is about $1{\cdot}5 \times 10^7$. For this to be produced the approximate coincidence of periods must last long enough, and also the damping due to internal viscosity must be small. Now the change of α in one period is $2\pi d\alpha/dw$, and hence the magnification, in the absence of viscosity, would be $2\pi\alpha \div$ the change of α in one period. Thus the change of α in a period must have been less than $4 \times 10^{-7}\alpha$. The period was about 2 hours, so that there would be time to develop a sufficient magnification if α did not vary by a large fraction of itself in 500 years. The damping effect of small viscosity on the free oscillations of a fluid globe is such as to make the amplitude of an oscillation of ellipsoidal type die down like $e^{-t/\tau}$, where*

$$\tau = \frac{a^2}{5\lambda} = \frac{2}{\mu},$$

a being the radius and λ the kinematic viscosity. The possible magnification is therefore $\alpha/\mu = a^2\alpha/10\lambda$. This gives

$$\lambda < 2{\cdot}4 \times 10^6 \text{ cm.}^2/\text{sec.}$$

This is comparable with the viscosity of shoemaker's wax found by Trouton and Andrews.

3·5. The distorted form of the earth would be very long and narrow. The dense interior, instead of being drawn towards the sun, would be depressed, in accordance with the results of 3·21. Thus when the elongation became so great that the mass became unstable and the end broke off, the detached portion would be at a considerable distance from the centre of the earth, certainly several times the undisturbed radius, and it would be composed chiefly of materials from the outer regions of the earth.

* Lamb, *Hydrodynamics*, 1924, 607 (after G. H. Darwin).

Its linear dimensions would be decidedly less than the radius of the earth, but of the same order of magnitude. The last two results agree with the actual size and density of the moon; it will be seen later (Chapter XIV) that the first has had an important influence on its history.

It should be noticed that, although the elongation would be towards the sun throughout the changes, the matter of the earth would always be rotating within the slowly moving surface, just as water can revolve within a fixed elliptical dish, each particle within the earth revolving about the axis in the period of rotation, whereas the surface would only complete its revolution in a year. Thus when rupture occurred the detached portion would have a considerable transverse velocity and therefore would not fall back into the earth.

CHAPTER IV

The Resisting Medium

"Friction produces heat." *Any School Physics.*

4·1. *Origin of the Medium.* So far we have seen that

(1) the disruption of the primitive sun by a passing star could have led to the formation of the planets;

(2) the tidal action of the sun on the outer planets, the first time they passed perihelion, may have led to the formation of satellites and of independent small planets;

(3) such planets as stayed with the sun, and such satellites as remained with their original primaries, would have had direct revolutions;

(4) the fact that the smallest planets are the densest is explicable on the same hypothesis;

(5) the fact that the planes of the motions of all these bodies are nearly coincident is similarly explicable;

(6) the moon may have been produced from the earth by the solar tides, magnified by resonance.

Several striking facts about our system, however, still remain unexplained. It is necessary to provide explanations of

(1) the smallness of the eccentricities of the orbits of the planets and satellites;

(2) the retrograde motions of two satellites of Jupiter, one of Saturn, four of Uranus, and one of Neptune;

(3) the curious numerical relations between the mean motions of several satellites;

(4) the retrograde rotations of Uranus and Neptune, and the direct rotations of the earth and Mars;

(5) the formation of asteroids;

(6) the acquirement by Mars of two small satellites;

(7) the recession to its present distance of the moon, which can have been only 10,000 to 20,000 km. from the centre of the earth when it was formed.

Some of these questions have been answered, and hitherto none has been proved unanswerable; but until all have been answered the theory of the origin and development of the solar system cannot be considered complete. The successes so far attained, however, are enough to encourage the cosmogonist to hope for the attainment of the others.

So far little explicit use of any form of friction has been made in the theory, although it has been virtually assumed in the supposition that any ejected body reabsorbed into its parent would become an integral part of it, the whole rotating as a rigid body. Friction must, however, have influenced the history of the solar system in at least two other ways, namely, by the action of a resisting medium and by tidal friction, of which the former has probably had the more far-reaching effects, although it happens that tidal friction has been exceptionally important in determining the evolution of the earth and moon in particular (see later, Chapter XIV).

It has already been indicated that the matter ejected from the sun would not all be included in the planets and their satellites. Much of it would be lost on account of the inadequacy of the gravitative power of the distended nuclei to retain their lighter constituents, and the thinner parts of the filament would probably be unable to condense at all, but would spread out at once. This lost matter would be dispersed throughout the system, and would form the resisting medium. It is clear from its formation that it would be largely or entirely gaseous. It would have the same origin as the planets, and therefore would have been deflected transversely by the star, just as the planets were. Hence every part of it would have a direct revolution about the sun from the very beginning. The parts would, on the other hand, revolve in widely different periods, and would undergo diffusion at the same time, until the whole system was filled with tenuous matter. Differences in the periods of revolution would make some parts move outwards while others meeting them were moving inwards, and thus the radial motions would be quickly annulled by turbulence and viscosity. Thus the resisting medium would be a gas, its parts revolving around the sun in the same direction as the planets, and describing approximately circular paths. We have no knowledge of the composition of the matter originally ejected that might enable us to estimate the mass of the medium or the distribution of density within it. This can be found, if at all, only from the effects that we assume it to have produced. It will be found, however, that it leads to another inference that is capable of independent test.

4·2. *Density Distribution and Motion of the Medium.* Let us now consider the nature of the motion of the medium. For the reasons already given, the medium will be supposed to be symmetrical about an axis, which will be taken to be the axis of z, and every part of it will be supposed to revolve uniformly in a circle about this axis. Taking rectangular coordinates x and y in fixed directions in the equatorial plane of the mass and through the centre of the sun, we put

$$x^2 + y^2 = \varpi^2 \qquad \text{.........(1),}$$

and denote the velocity at any point by $\omega\varpi$ perpendicular to the meridian plane. So far ω is unspecified and need not be a constant. Then the equations of motion of the medium are

$$\left.\begin{aligned}\frac{\partial U}{\partial x}-\frac{1}{\rho}\frac{\partial p}{\partial x}&=-\omega^2 x\\ \frac{\partial U}{\partial y}-\frac{1}{\rho}\frac{\partial p}{\partial y}&=-\omega^2 y\\ \frac{\partial U}{\partial z}-\frac{1}{\rho}\frac{\partial p}{\partial z}&=0\end{aligned}\right\} \qquad \ldots\ldots\ldots(2),$$

where U is the gravitation potential, p the pressure, and ρ the density. In differential notation these may be written

$$dU-\frac{1}{\rho}dp=-\tfrac{1}{2}\omega^2 d\varpi^2 \qquad \ldots\ldots\ldots(3).$$

If V denote the absolute temperature, we have

$$p=R\rho V \qquad \ldots\ldots\ldots(4),$$

where R is a constant if the composition of the material is uniform. Also

$$U=fM/r \qquad \ldots\ldots\ldots(5),$$

where f is the constant of gravitation, M the mass of the sun, and

$$r^2=x^2+y^2+z^2 \qquad \ldots\ldots\ldots(6).$$

The mass of the medium being supposed a small fraction of that of the sun, its gravitation may be neglected.

The temperature in an approximately steady state is determined by the condition that each part of the mass receives just as much heat as it radiates. The other parts of the envelope are by hypothesis much colder than the sun, and we may assume that each part is warmed wholly by solar radiation, and radiates its heat away at a rate proportional to V^4. The rate of receipt of heat is proportional to $1/r^2$, and therefore

$$V=ar^{-\frac{1}{2}} \qquad \ldots\ldots\ldots(7),$$

where a is some constant. Then equation (3) gives, after substitution from (4), (5) and (7),

$$\left(-\frac{fM}{ar^{\frac{3}{2}}}+\frac{1}{2}\frac{R}{r}\right)dr-R\frac{d\rho}{\rho}=-\frac{1}{2}\frac{\omega^2 r^{\frac{1}{2}}}{a}d\varpi^2 \qquad \ldots\ldots\ldots(8).$$

It follows that $\omega^2 r^{\frac{1}{2}} d\varpi^2$ is a perfect differential, and

$$\omega^2=r^{-\frac{1}{2}}F^2 \qquad \ldots\ldots\ldots(9),$$

where F is a function of ϖ, as yet unspecified. Substituting in (8) and integrating, we have

$$R\log\rho=\frac{2fM}{ar^{\frac{1}{2}}}+\tfrac{1}{2}R\log r+\frac{1}{2a}\int F^2 d\varpi^2+\text{const.} \ldots\ldots\ldots(10).$$

Accordingly, since the integrand in the last variable term on the right is essentially positive, the density when r is great enough must increase with r at least as fast as $r^{\frac{1}{2}}$, and therefore the whole mass must be infinite. Hence an exact steady motion is impossible with the conditions specified; the fluid must necessarily flow outwards to some extent. The loss will, however, not affect the distribution of density appreciably unless $fM/aRr^{\frac{1}{2}}$ is small. Supposing, as is reasonable, that the temperature 10^{13} cm. from the sun (roughly the distance of Venus) was 300° abs., (7) shows that a is 10^9 c.g.s. Cent. units, and if the medium be supposed to consist of hydrogen, R is 4×10^7 c.g.s. Then the ratio in question becomes equal to unity when r is 4×10^{19} cm. Thus outward flow would not be important in masses of the size of the solar system, and in them the second term on the right of (10) and the term in R/r in (8) are unimportant.

Suppose now that on the equatorial plane of the medium

$$\omega^2 = \lambda^2 fM\varpi^{-3} \qquad \text{.........(11)},$$

so that the velocity at any point is λ times what it would be if every part of the medium were describing a circle freely under gravity. Then (8) gives on the equatorial plane

$$R\frac{d\rho}{\rho} = \frac{fM}{ar^{\frac{3}{2}}}(\lambda^2 - 1)\,dr \qquad \text{.........(12)}.$$

Thus if the angular velocity at any point exceeds the circular velocity, the density will increase outwards, while if it is less than the circular velocity, the density will increase inwards. The more closely the velocity approximates to the circular velocity, the more nearly will the density be uniform. To indicate the importance of this approximation, let us consider the special case of no rotation, supposing the density where

$$r = 2 \times 10^{12}\ \text{cm.} \qquad \text{.........(13)},$$

to be equal to 5×10^{-5} gm./cm.³ This density would imply a mass equal to that of the sun spread out a third of the distance to Mercury. Then with the data already adopted, the density at any point is given by

$$\log\frac{\rho}{5 \times 10^{-5}} = -\frac{1{\cdot}2 \times 10^{10}}{r^{\frac{1}{2}}} \qquad \text{.........(14)},$$

and the density near the orbit of Mars is therefore less than 10^{-1000}. Beyond the orbit of Mars it would be still less, and therefore the mass of the medium between the orbits of Mars and Neptune, the radius of Neptune's orbit being 5×10^{14} cm., is less than $\frac{4}{3}\pi . 10^{-1000}(5 \times 10^{14})^3$ grams, a quite inappreciable fraction of a gram, and totally incapable of producing a noticeable influence on the orbit of the smallest asteroid. Hence the resisting medium could not be of any cosmogonical importance unless each part of it revolved with very nearly the velocity appropriate to a planet

moving in a circular orbit at the same distance. With very slight departures from this relation an extremely wide range of variation of density within the medium could be realized.

The last point requires considerable emphasis, since most writers that have made use of a resisting medium in cosmogonical theories have assumed without investigation that such a medium would be at rest. A medium at rest would oppose a steady resistance to the motion of a planet, and would thereby reduce its total energy and make it fall towards the sun. Thus a stationary resisting medium would cause all planets to approach the sun, and this result has been habitually assumed in discussions of the effects of such a medium. From what has just been shown it appears that the medium would move in such a way that a planet in a circular orbit would have no motion relative to it, and therefore would experience no resistance and be quite unaffected by it.

4·3. *Effect of a Planet on the Medium.* In the case of a small satellite moving in a variational orbit around a planet and the sun, the corresponding proposition would be that the satellite would still always be moving in the same way as the medium around it; so that every part of the medium would have to move in a variational orbit, fluid pressure not affecting its motion in the least. This will be shown to be inconsistent with the equation of continuity, unless the temperature is the absolute zero.

4·31. It appears that the motion of a medium around a planet and the sun, the planet's orbit being circular, could not be even roughly steady. It can be shown that the only possible steady motions of a gaseous medium or a swarm of meteors are such that the whole medium is statistically rotating with the planet like a rigid body. For, let us take the axes of x and y through the sun, the axis of x being always towards the planet. If the mean motion of the planet be n and the gravitation potential U, the equations of motion of any particle are

$$\left.\begin{aligned}\frac{d^2x}{dt^2} - 2n\frac{dy}{dt} - n^2x &= \frac{\partial U}{\partial x}\\ \frac{d^2y}{dt^2} + 2n\frac{dx}{dt} - n^2y &= \frac{\partial U}{\partial y}\\ \frac{d^2z}{dt^2} &= \frac{\partial U}{\partial z}\end{aligned}\right\} \quad \text{.........(1).}$$

Multiplying by $\frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt}$ and adding, we obtain on integration the equation

$$\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2 - n^2(x^2+y^2) = 2U + C \quad \text{.........(2),}$$

where C is a constant throughout the motion of the particle. This is

Jacobi's integral, and it is known that the equations of motion possess no other first integral.

Now consider a swarm of particles, which may be solid bodies or the molecules of a gas, revolving about the sun and planet. Let the number in the element of volume $dx\,dy\,dz$, the rates of change of whose coordinates are between u and $u + du$, v and $v + dv$, and w and $w + dw$ respectively, be $\phi\,(uvwxyz)\,du\,dv\,dw\,dx\,dy\,dz$. (It is to be noticed that u, v, w are to be regarded as rates of change of the coordinates and not as velocity components referred to fixed axes momentarily coincident with the moving ones.) In the absence of collisions each particle will move in time dt to a position $x'y'z'$, and will be changing its coordinates at rates u', v', w', where to the first order in dt

$$x' = x + u\,dt, \text{ etc.,} \quad u' = u + \frac{du}{dt}\,dt, \text{ etc.} \qquad \text{.........(3).}$$

But $\dfrac{du}{dt}$, etc., are known functions of the coordinates and their rates of change, so that x', y', z', u', v', w' are known as functions of x, y, z, u, v, w and of dt. All particles in the original element move in time dt into the new element, and therefore

$$\phi'\,(u'v'w'x'y'z')\,du'\,dv'\,dw'\,dx'\,dy'\,dz' = \phi\,(uvwxyz)\,du\,dv\,dw\,dx\,dy\,dz \text{...(4),}$$

where ϕ' indicates that the velocity and density distribution after time dt is being considered. If the distribution is to be steady, so that the number of particles in a given element as regards position and velocity will be the same for all time, ϕ' is the same as ϕ. Now

$$\frac{du'\,dv'\,dw'\,dx'\,dy'\,dz'}{du\,dv\,dw\,dx\,dy\,dz} = \frac{\partial\,(u'v'w'x'y'z')}{\partial\,(uvwxyz)}$$

$$= \begin{vmatrix} 1 & 0 & 0 & \dfrac{\partial}{\partial x}\dfrac{du}{dt}\,dt & \dfrac{\partial}{\partial y}\dfrac{du}{dt}\,dt & \dfrac{\partial}{\partial z}\dfrac{du}{dt}\,dt \\ 0 & 1 & 0 & \dfrac{\partial}{\partial x}\dfrac{dv}{dt}\,dt & \dfrac{\partial}{\partial y}\dfrac{dv}{dt}\,dt & \dfrac{\partial}{\partial z}\dfrac{dv}{dt}\,dt \\ 0 & 0 & 1 & \dfrac{\partial}{\partial x}\dfrac{dw}{dt}\,dt & \dfrac{\partial}{\partial y}\dfrac{dw}{dt}\,dt & \dfrac{\partial}{\partial z}\dfrac{dw}{dt}\,dt \\ dt & 0 & 0 & 1 & 0 & 0 \\ 0 & dt & 0 & 0 & 1 & 0 \\ 0 & 0 & dt & 0 & 0 & 1 \end{vmatrix}$$

$$= 1 + \text{terms in } dt^2 \qquad \text{.........(5).}$$

It is proved in works on the dynamical theory of gases that collisions do not affect this result (cf. Jeans, *Dynamical Theory of Gases*, 2nd edition, p. 226).

Substituting in (4) we see that

$$\phi\,(u'v'w'x'y'z') = \phi\,(uvwxyz) \qquad \text{.........(6)}$$

to the first order in dt; so that $\quad \dfrac{d\phi}{dt} = 0$ ....(7),

if u', v', w', x', y', z' are related to u, v, w, x, y, z according to the relations (3), with the values of $\dfrac{du}{dt}, \dfrac{dv}{dt}, \dfrac{dw}{dt}$ given by (1). In other words $\phi =$ constant is a first integral of the equations of motion. Further, since by hypothesis the motion is steady, ϕ does not involve the time explicitly. Thus this theorem, proved by Jeans for fixed axes, is readily extended to moving axes.

In the case we are considering there is no first integral except the Jacobi integral. It follows that in a steady state ϕ must be a function of

$$u^2 + v^2 + w^2 - n^2 (x^2 + y^2) - 2U = u^2 + v^2 + w^2 - \nu, \text{ say} \quad \text{...(8).}$$

Put then
$$\phi = \chi (\nu - u^2 - v^2 - w^2) \quad \text{.........(9).}$$

The numerical density λ can be found by putting

$$u = \xi \nu^{\frac{1}{2}}, \quad v = \eta \nu^{\frac{1}{2}}, \quad w = \zeta \nu^{\frac{1}{2}} \quad \text{.........(10),}$$

when
$$\lambda = \iiint \phi \, du \, dv \, dw$$
$$= \iiint \chi \{\nu (1 - \xi^2 - \eta^2 - \zeta^2)\} \nu^{\frac{3}{2}} d\xi \, d\eta \, d\zeta \quad \text{.........(11),}$$

the limits being $-\infty$ to ∞ in all cases. Hence λ, and therefore ρ, are functions of ν alone. Since ϕ is an even function of u, v, w, we see further that the medium as a whole has no systematic motion with reference to the axes. Both of these results are independent of whether the medium consists of a gas or of solid particles.

4·32. The first of these results finds a simple application in connexion with the Moulton-Gyldèn theory of the Counterglow*. According to this theory the counterglow is caused by light reflected from particles describing orbits about the sun and earth jointly, a particularly large number of which are visible at any time in the part of the sky directly opposite to the sun. From the result that the density is a function of ν alone we infer that the illumination will be greatest where the observer is looking through the greatest depth between consecutive surfaces of the system $\nu =$ constant. But just opposite to the sun is a place where one of these surfaces has a conical point, and it is at this point that the distance between surfaces of the system is greatest, just as a hyperbola is furthest from its asymptotes near the centre. Hence at this point the observer is looking through the greatest depth of matter, and therefore sees a patch of reflected light.

4·33. Next, consider the motion of a small particle through the medium. The resistance to its motion will be opposite to the direction of its velocity relative to the medium, and will vanish with the relative velocity. But the medium has no systematic motion with regard to the

* *Bulletin Astronomique*, t. 1; *Astronomical Journal*, No. 483.

axes. If then the coordinates of the particle are x, y, z, the components of the retardation will be $-\kappa\dot{x}$, $-\kappa\dot{y}$, $-\kappa\dot{z}$, where κ is positive. Hence its equations of motion are

$$\left.\begin{aligned}
\frac{d^2x}{dt^2} - 2n\frac{dy}{dt} - n^2x &= \frac{\partial U}{\partial x} - \kappa\frac{dx}{dt}\\
\frac{d^2y}{dt^2} + 2n\frac{dx}{dt} - n^2y &= \frac{\partial U}{\partial y} - \kappa\frac{dy}{dt}\\
\frac{d^2z}{dt^2} \qquad\qquad\qquad &= \frac{\partial U}{\partial z} - \kappa\frac{dz}{dt}
\end{aligned}\right\} \quad \text{.........(1).}$$

Multiply by $\frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt}$ respectively and add. Then

$$\frac{d}{dt}\left\{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2 - n^2(x^2+y^2) - 2U\right\} = -2\kappa\left\{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2\right\}$$

$$\text{.........(2).}$$

Hence the function $\nu - \dot{x}^2 - \dot{y}^2 - \dot{z}^2$ steadily increases with the time. Thus if we denote it by μ, we see that ν is always greater than μ, which is itself in general steadily increasing at a finite rate. Hence in time μ, and therefore ν, will exceed any finite limit. Whatever be the value of μ at any one time, the motion will be such that at any subsequent time the particle will be at a place where ν is greater than that value of μ. Now the surfaces $\nu =$ constant are closed, and the greater ν is the more closely they approach the three places where ν becomes infinite, situated at the sun, the planet, and at an infinite distance. Thus the motion of the third body will become more and more restricted until it is ultimately forced to revolve as an independent planet, as a satellite of the planet, or is expelled from the system; in either of the two former events it will steadily approach its primary. The only exceptional case is that where the coordinates of the third body remain the same permanently, so that $\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2$ is zero. In such a case friction will not affect the motion. The only stable motion of this type is that where the three bodies are at the corners of an equilateral triangle, and in this case ν is a minimum. Hence if any small displacement from the equilateral triangle position takes place, the disturbance will steadily increase until the third body becomes either an independent planet or a satellite. We notice also that a small body with a greater mean distance than the planet must approach the sun and hence become either an interior planet or a satellite. It thus appears possible that the capture of satellites can occur if the resisting medium is in a state of steady motion.

The above argument, however, cannot be applied entirely to the actual evolution of the solar system. In a gaseous medium ϕ must be of the form $Ne^{h^2(\nu - u^2 - v^2 - w^2)}$, where N and h are constants. It follows that the density is proportional to $e^{h^2\nu}$. It can hence be easily shown that the results

based on 4·2 (12) are not appreciably altered in the regions remote from the planet. In the neighbourhood of Jupiter, again, the variation of the temperature would not be considerable, and we should have nearly

$$\log \rho = \frac{\nu}{RV}.$$

Taking $V = 100°$, we see that the difference between the values of $\log \rho$ at the surface of Jupiter and 100 radii away is of the order of 5000, indicating, as in 4·2, that the assumption that the more remote satellites of Jupiter have been appreciably affected by the resisting medium is inconsistent with the assumptions of a reasonable density near Jupiter and of steady motion of the medium. It appears, therefore, that the medium cannot have been even locally in a state of steady motion, and that the results obtained on this hypothesis, though of interest as suggesting possible modes of development, cannot be regarded as giving the correct theory of the influence of a resisting medium on satellites.

4·34. We may, however, recur to the hypothesis of 4·3 that the medium at any point was moving like a planet revolving in a periodic orbit through the point. This velocity is fixed for given coordinates, and therefore the motion would be a steady one: but for any other temperature than the absolute zero we have found the only steady motion dynamically possible, and it is not this one. Thus a steady motion of this type is possible only if the temperature is the absolute zero. But the temperature cannot be the absolute zero. Hence this type of motion is impossible, and the medium must exert a dissipative influence on the third body. Some secular effect on the mean distance of a satellite from its primary is therefore probable, though not certain.

4·4. *Effect of the Resisting Medium on the Mean Distances of Satellites.* Now considering the solar system as it is, we notice several striking facts that are readily explicable if it is true that a resisting medium would make satellites approach their primaries, and might even enable a planet to capture a body previously moving as an independent planet and force it to move as a satellite of its own. The retrograde satellites of Jupiter and Saturn are readily explained on the hypothesis that they were produced by the tidal action of the sun on the planets, but left their parents at once and became independent planets; and that they afterwards were captured by their present primaries, which may or may not have been their original parents. The same may also apply to some direct satellites, especially to the two satellites of Mars, the sixth and seventh of Jupiter, and Iapetus. Such an abnormal origin is suggested in the case of the satellites of Mars by their small size; such small bodies could not have been formed from a planet such as Mars, which must have liquefied almost immediately. For J VI, J VII, and Iapetus it is suggested by the fact that whereas

the inner satellites of Jupiter and Saturn are regularly spaced as regards distance from the primary, wide gaps separate them from the orbits of these outer ones. It is possible, however, that the gap is due to the density of the medium at the distance of these satellites having been very small in comparison with that nearer the primaries, so that the inner satellites were led to approach their primaries, leaving the outer ones almost unaffected.

4·41. The hypothesis that a resisting medium in some cases did make satellites approach their primaries is strongly confirmed by the existence of Saturn's rings. It is generally believed that the rings represent the fragments of a solid satellite that was broken up by the tidal action of Saturn through being too near to the primary for the mutual attractions of its parts to resist the disruptive tendency. Now if the rings were initially at their present distance, they could never have condensed; for the efficacy of the tidal action in comparison with the mutual gravitation of the parts would increase with distension, and therefore if the former were able to disrupt it when solid, still more capable would it be when gaseous. Hence the mass would have remained gaseous until it diffused away. Hence the satellite that formed the rings must have been formed when beyond the danger zone and have afterwards been brought within it by some disturbing influence. No other agency than a resisting medium has been suggested that could produce such an approach.

4·42. Several curious numerical relations hold between the mean motions of various satellites. The mean motions n_1, n_2, n_3 of J I, J II, and J III are approximately in the ratios 4 : 2 : 1, while $n_1 - 2n_2$ and $n_2 - 2n_3$ are exactly equal to each other and approximately to $\frac{1}{300}n_1$.

The mean motions of Mimas, Enceladus, Tethys, and Dione are nearly in the ratios 6 : 4 : 3 : 2, none departing from the values corresponding to these ratios by more than 3 per cent. Those of Titan and Hyperion are as 4·004 : 3. On any theory yet advanced of the origin of satellites it is very difficult to see how such relations could have subsisted from the beginning; but if the mean distances of the satellites were varying continually owing to a resisting medium, such ratios would occur several times during the evolution, and if the corresponding states of the systems were stable they would thenceforth persist.

4·43. The satellites of Uranus and Neptune were probably formed by tidal disruption of their primaries by the sun. Any satellite formed in this way would be direct; but if the axis of rotation of the primary was strongly inclined to the ecliptic, the ellipticity of figure of the primary would make the plane of the satellite's orbit revolve; and if a resisting medium were available to damp down the component of the motion of the satellite parallel to the axis of the planet, the satellite would come to revolve in

the plane of the equator of the primary, even though the primary might have a retrograde rotation.

4·5. *Evolution of the Medium, and its Effect on Mercury.* Let us now consider the manner of evolution of the resisting medium. So far its internal viscosity, diffusion, and thermal conductivity have all been ignored. Since viscosity must necessarily produce a secular effect on the motion of any mass that is not moving either irrotationally or with the same rotation at all points, neither of which conditions is satisfied by a medium moving as this one would, the motion of the medium must undergo a slow and steady change. The nature of the change is easily seen. The fast-moving interior will tend to drag forward the slower-moving exterior, and thus will increase its energy and make it recede from the sun. Thus the outer parts will be slowly expelled from the system. The inner parts, on the other hand, will have their motion delayed, and will therefore gradually fall into the sun. In time, therefore, the resisting medium will cease to exist. Diffusion would not affect the behaviour of a medium consisting of only one material; thermal conductivity would be continually transferring heat from the inside to the outside, but this would probably produce only a slight permanent change in the temperature distribution, and not a secular degeneration. If now ρ, μ, l, and τ be of the order of magnitude of the density, true viscosity, linear dimensions, and time of degeneration through viscosity, we have

$$\tau = \frac{\rho l^2}{\mu} \qquad \text{.........(1)},$$

from an analogy with the behaviour of other systems changing on account of viscosity, or from considerations of dynamical similarity. Considering the motion of the medium within the orbit of Mercury, we can take

$$l = 6 \times 10^{12} \text{ cm.} \qquad \text{.........(2)},$$

$$\mu = 10^{-4} \text{ gm./cm. sec.} \qquad \text{.........(3)}$$

(since μ is independent of the density), and therefore

$$\tau \text{ is of order } 4 \times 10^{29}\rho \qquad \text{.........(4)}.$$

Consider next the motion of the planet Mercury, supposing it to have been moving in a highly eccentric orbit. The angular velocity about the sun of the matter near the planet's orbit being n, and the velocity of the planet relative to the medium being Υ, the resistance to the motion of the planet is $\frac{1}{2}\pi\rho a^2\Upsilon^2$, where a is the radius of the planet*, since the relative velocity is much greater than the velocity of sound in the medium. Hence the time needed to reduce the relative motion to $1/e$ of its initial amplitude is of order $m/\frac{1}{2}\pi\rho a^2\Upsilon$, where m is the mass of the planet. Taking

$$m = 2 \times 10^{26} \text{ gm.}, \quad a = 2{\cdot}6 \times 10^{8} \text{ cm.}, \quad \Upsilon = 2 \times 10^{6} \text{ cm./sec.},$$

this time is $$1000/\rho \qquad \text{.........(5)}.$$

* F. A. Lindemann and G. M. B. Dobson, *Proc. Roy. Soc.* **102 A, 1922, 413.**

If the whole extent of the medium was greater than, but of the same order of magnitude as, the distance of Mercury from the sun, the results of the last two paragraphs could be combined to give an estimate of the age of the solar system. For, if the time needed by the medium to disappear was short in comparison with the time required to produce a considerable effect on the eccentricity of the orbit of a planet, the medium would have gone before the eccentricity had been appreciably reduced, and the eccentricity would still be great. On the other hand, if the medium lasted much longer than the time required to affect the eccentricity considerably, the eccentricity would have been reduced practically to zero instead of only to about $\frac{1}{4}$. Hence these two times must be comparable. Supposing them to be equal, we find that the density must have been of order 5×10^{-14} gm./cm.^{3}, and the time taken of order 2×10^{16} seconds or 6×10^8 years. On the other hand the actual medium must have had a much wider extent, and therefore must have been acting upon Mercury during the time required for a much larger medium to degenerate. If for instance matter from the distance of the orbit of Jupiter—to take what is perhaps an extreme hypothesis—passed across the orbit of Mercury and was ultimately absorbed into the sun, l in (1) would have to be taken equal to 8×10^{13} instead of 6×10^{12} cm., and thus τ would be of order $5 \times 10^{31}\rho$. Combining this with (5) we find that the time required would be of order 7×10^9 years, and the density would be 4×10^{-15} gm./cm.^{3}

The coefficient of true viscosity is independent of the density, and it has been shown that so long as the medium was affecting the planets, and we can suppose its motion dominated by the sun, its motion would remain the same. Hence however much the density declined, so long as the medium behaved like a gas, the rate of communication of angular momentum across any sphere within the fluid would remain the same; therefore the rate of absorption of matter into the sun or expulsion of matter from the system would remain the same. Hence in a further duration, comparable with that required for the medium to have its density reduced to half what it was initially, practically the whole of it would have disappeared. What remained would indeed have to be of density so low that the gas laws would not apply.

The only gaseous matter of sufficient density to be observable that exists outside the planetary atmospheres is that which reflects the zodiacal light, and it is natural to suppose that this is the last relic of the resisting medium*. Its density, estimated from its luminosity†, is of order 10^{-18} gm./cm.^{3} Now the length of the mean free path of a hydrogen

* Many astronomers are favourable or neutral to the theory that the zodiacal matter consists of solid meteors. If so its life would be of the order of 10^7 years (*M.N.R.A.S.* 77, 1916, 93–94). The light is however polarized as if scattered by the molecules of a gas (H. N. Russell, R. S. Dugan, and J. Q. Stewart, *Astronomy*, 358–360).

† Jeffreys, *M.N.R.A.S.* 80, 1919, 139.

molecule at normal temperature and pressure is $1{\cdot}83 \times 10^{-5}$ cm., and in other circumstances is inversely proportional to the density. Hence in the zodiacal matter the mean free path is of order 10^9 cm., much less than the radial extent of this matter. Each molecule must therefore experience many collisions in every revolution around the sun, and this is the condition that the gas laws may apply. Thus the degeneration of the zodiacal matter must still be going on, and therefore it is probable that the whole age of the system is not more than twice the time needed to reduce the eccentricity of the orbit of Mercury to $1/e$ of what it was at the commencement. Hence the tidal theory of the origin of the solar system suggests that the age of the system is of order 10^9 to 10^{10} years. There are of course many sources of error in the data, the chief being in the primitive distribution of mass, but it is interesting to notice that the age obtained is of the same order of magnitude as the age of the earth inferred from the phenomena of radioactivity, which will be discussed later.

4·51. On the hypothesis that the density of the medium was 5×10^{-14} gm./cm.^{3}, the mass within the orbit of Mercury would be about 4×10^{25} grams, decidedly less than that of the planet. On the hypothesis that it was 4×10^{-15} gm./cm.^{3}, the mass within the orbit of Jupiter would be about 8×10^{27}, rather more than the mass of the earth, which is reasonable. If the density was 4×10^{-15} gm./cm.^{3} throughout the system, the mass within the orbit of Neptune would be comparable with that of Jupiter, which again is reasonable.

4·6. *Effects on other Planets*. It would be of interest to test the theory by application to the eccentricities of the orbits of the other planets, but unfortunately one meets with a serious mathematical difficulty. The gravitation of each planet would cause a condensation in the medium around it. The formula of Lindemann and Dobson, for the resistance of a gas to a body moving with a relative velocity much greater than that of sound, depends essentially on the condition that the matter around the body, right up to its surface, is continually changing. In the case of a gravitating planet this condition would hold only if the relative velocity of the gas towards it was greater than the velocity of escape from the surface of the planet; otherwise a portion of the gas would be retained by the gravitation of the planet. Thus a permanent gaseous condensation would be formed around the planet, and would be forced through the medium by it, so that the effective resistance to the motion of the planet would not be determined by the surface of the planet itself, but by that of this condensation. It seems reasonable to conjecture that the effective radius of such a condensation would be such that the velocity of a particle moving in a parabolic orbit about the planet at that distance (the disturbance due to the sun being ignored) is comparable with the velocity of the medium relative to the

planet. If b denote this radius, m the mass of the planet, M that of the sun, and if the velocity of the planet relative to the medium is λ times the velocity rn of the medium itself, this gives

$$\left(\frac{2fm}{b}\right) = O(\lambda rn)^2 = O\left(\lambda^2 \frac{fM}{r}\right) \qquad \text{.........(1)},$$

whence

$$b = O\left(\frac{2mr}{\lambda^2 M}\right) \qquad \text{.........(2)}.$$

Taking as a preliminary standard the impossible case where the planet is at rest and therefore λ equal to 1, we find for the various planets the following values of b, in kilometres: Mercury 13; Venus 500; Earth 1000; Mars 130; Jupiter $1{\cdot}6 \times 10^6$; Saturn 8×10^5; Uranus $2{\cdot}5 \times 10^5$; Neptune $4{\cdot}4 \times 10^5$.

For the four terrestrial planets these numbers are much smaller than the actual radii, and the neglect of gravitation would therefore be justified if λ was equal to unity; thus in the early stages of their careers, when the eccentricities of their orbits were great, there would be little gravitational condensation around them, and the effective surface would be practically the solid surface. But when the eccentricities became small, the value of λ would sink to something between the eccentricity and twice the eccentricity, and when this fact is allowed for, the only planet for which b is still less than the actual radius is Mercury. Hence the method of estimating the ages of the solar system already employed is unsuitable for any planet except Mercury. The values of b for the four great planets have always been greater than the radii.

4·61. For a non-gravitating planet the time required to produce a given change in the eccentricity is, by 4·5, proportional to $mr^{\frac{1}{2}}/\rho a^2 M$. For a great planet we may write the effective radius b for a. For Mercury $mr^{\frac{1}{2}}/a^2M$ is $1{\cdot}2 \times 10^{-10}$ km.$^{-\frac{3}{2}}$; for Jupiter $mr^{\frac{1}{2}}/b^2M$ is $1{\cdot}1 \times 10^{-11}$ km.$^{-\frac{3}{2}}$. Thus the effect of the gravitational condensation is so great that if the densities were the same and the eccentricity of the orbit of Jupiter large, the resistance would reduce the eccentricity of the orbit of Jupiter nine times as fast as that of Mercury. For smaller eccentricities a for Mercury would remain the same, while b for Jupiter would increase further. It would be dangerous to proceed far without more knowledge than we possess about the distribution of density in the resisting medium, but this effect of gravitational condensation offers at least a very striking suggestion as to the reason why the outer planets have small eccentricities, while Mercury has the largest in the system.

4·62. Another fact that is possibly related to the last is that the value of b for Jupiter, with λ equal to 1, is nearly the mean distance of the fourth satellite; for Saturn, about the distance of Titan; for Uranus, between those of Titania and Umbriel; and for Neptune, approximately

the distance of its single satellite. Thus all the satellites except J VI, J VII, Iapetus, and the retrograde ones of Jupiter and Saturn would be within the gravitational condensations around their primaries almost from the start, and therefore their motions would have undergone the greatest disturbances from the resistance. The effects on the outer satellites would perhaps not become great until the reductions in the eccentricities of their primaries had considerably increased the sizes of the condensations. This may afford an alternative reason to capture for the wide gaps between the orbits of J IV and J VI, and between those of Hyperion and Iapetus.

4·7. So far no attempt has been made to account for the asteroids, which still offer an outstanding difficulty in this theory, as in every other. Their small size in comparison with Mercury indicates that, if they existed while the resisting medium was still exerting an appreciable influence on the terrestrial planets, their eccentricities must have been reduced to zero. Thus they must have been formed after the medium had almost disappeared, and therefore are not lost satellites. The fact that none of their mean distances is much less than that of Mars or greater than that of Jupiter suggests that they are fundamentally related in origin; and the most natural explanation is that they were formed by the disruption of a primitive planet. Their total mass can hardly exceed a hundredth of that of the earth; this is much less than that of Mercury, but comparable with those of the great satellites of Jupiter, and the possibility that the asteroids were formed from a primitive planet which was itself a lost satellite of Jupiter may be entertained. The possible modes of rupture include explosion, rotational instability, and the tidal action of the sun or Jupiter. Rotational instability would only give a planet with one satellite comparable in size with itself; the same applies to the tidal action of the sun, for this would have to be magnified by resonance to lead to rupture, and then the theory that has been used to account for the origin of the moon would be applicable. Explosion is possible, though there is little evidence for it. If the planet contained enough radioactive or chemically active matter to heat part of its interior up to the boiling point, the disruptive stresses might come to exceed the small gravitative power of such a mass and permit explosion, and if one explosion took place there is no reason why others should not follow, since gravity would diminish at every rupture. Tidal disruption by close approach to Jupiter is also possible; the small planet would have to approach very close to the surface of Jupiter, and might be repeatedly broken up during a single encounter. The relative velocities of the fragments would be comparable with that which would have enabled a particle at the surface of the small planet to escape from its influence, which is about 2 km./sec. The average departure of the orbital velocities of the asteroids from the mean of all is about 3 km./sec. The relative velocities might be in any direction, for the small

planet might pass Jupiter considerably to the north or south of its orbital plane, and thus the orbits of the fragments might be considerably inclined, the possible inclinations again being of the same order as the actual ones. The aphelion distances of all would be almost equal to the distance of Jupiter when the encounter occurred.

The subsequent history of these bodies would be determined by such traces of the resisting medium as remained and by planetary perturbations. Their large eccentricities and inclinations agree equally well with the theories of explosion and tidal disruption. Some may have been captured; in particular it is possible that the satellites of Mars are captured asteroids. Large variations in the eccentricities of the orbits would be set up by the perturbations due to the planets, especially Jupiter, and the positions of the nodes and the apses would be continually varying. Thus the orbits would probably become considerably modified from their original form and position, and thus the fact that the smallest are contained wholly within the largest is consistent with the theory.

4·8. *Historical.* The investigation of the effects of a resisting medium goes back to Euler* and Laplace. They and other investigators before Chamberlin and Moulton, in 1901 to 1906, considered only media at rest relative either to the sun or to the average of the stars. The planetesimal system of Chamberlin and Moulton introduced virtually a resisting medium composed of solid particles in direct orbits. The earlier types of medium are open to the objection of 4·2, the latter to those of Appendix A; but the notion of a gaseous medium, revolving in the way specified here, introduced by me in 1918, seems to be a satisfactory compromise, combining the merits of the earlier theories while avoiding their more obvious faults. See in 1910† returned to a medium composed of cosmical dust, while using the older theories of the effects of a medium fixed with regard to the sun. A medium with these properties would immediately fall into the sun; but at the same time See introduced the notion of capture, both of satellites and planets. The latter cannot now be accepted, but we have seen that a notable number of facts would be explained if several satellites have been captured. See's arguments are always obscure and often definitely erroneous, owing largely to his incorrect views on the motion of the medium, especially near a planet. Nevertheless the notion of the capture of satellites is a positive contribution of value, though large gaps remain in the theory.

* *Phil. Trans.* **46**, 1749, 204.

† *Researches in the Evolution of Stellar Systems*, vol. 2.

CHAPTER V

The Age of the Earth

"Is there any thing whereof it may be said, See, this is new? it hath been already of old time, which was before us." Eccles. i. 10.

5·1. Several methods of estimating the age of the solar system in general, and of the earth in particular, have been suggested. The very plausible hypothesis that the eccentricity of the orbit of Mercury has been reduced to its present moderate value by the action of a resisting medium has been utilized for this purpose in 4·5, and indicates that the age of the system is probably between 10^9 and 10^{10} years. The age thus found is the time since Mercury first took shape as a planet, probably a few years at most from the ejection from the sun of the matter that formed it, and perhaps before the disturbing star had made its closest approach to the sun. Several other methods have been suggested for determining various long intervals in the earth's history, but the intervals determined are not in all cases the same, and a little attention must be given to the probable extent of the differences between them arising simply from the fact that they are not all measured from the same event. It will be seen that what we do in estimating a long interval of time is to consider some change that has taken place according to a known law; if we know both the law and the extent of the change between two definite events, we can calculate the time that elapsed between them. In the present problem the later event is in every case the present time; the earlier depends on the process considered. We have seen that the earth, like all other planets, was probably initially fluid. It cooled to the solid state by radiation from the surface, and even after solidification a further time would elapse before the surface became cool enough for water to condense on it. At a still later epoch denudation and redeposition formed the first sedimentary rocks. At some stage during this process the moon was formed. An estimate of the time that has elapsed since any one of these events will give information about the time since any other, when we have some knowledge of the intervals between these early events in the earth's history.

5·2. The chief methods (in addition to the one based on the eccentricity of the orbit of Mercury) that have been suggested for the estimation of the various intervals called 'the age of the earth' are as follows:

1. The age of an igneous rock can often be found directly by means of the ratio of the quantities of Uranium and Lead in it, the rate of degeneration of Uranium to Lead following a known law. This is available for rocks of a very great geological age, but these rocks are intrusive into

still older sedimentary rocks, and therefore the ocean must be still older than the oldest rocks whose ages have been determined in this way.

2. The age of the ocean could be found directly if we knew the total amount of salt in it and the rate of transfer of salt to the sea by rivers.

3. We could similarly find the age of the ocean if we knew the total quantity of sedimentary rocks on the earth's surface and the rate of disintegration of igneous rocks.

4. If, as seems probable, the earth's surface has been maintained at nearly the same temperature throughout geological time, we can show that the sun must have been radiating energy at almost its present rate throughout that time. If we can find the total amount of energy the sun has radiated away, we can find an upper limit to the time it can have been radiating at its present rate, which gives an upper limit to the time needed to form all the rocks known to geologists.

5. The time since the solidification of the earth may be found if we know its law of cooling and certain facts about the initial and present distributions of temperature.

6. Tidal friction has probably increased the period of the earth's rotation, from the period of 4 hours mentioned in 3·2, to the present period of 24 hours. If we knew its rate we could find the time since the birth of the moon.

5·3. *Radioactivity*. By far the most satisfactory of these methods appears to be the first. Its history dates from the discovery by Becquerel, in 1896, that uranium salts gave out rays capable of producing an effect on a photographic plate enclosed in opaque paper. This effect was found to be independent of the physical and chemical states of the uranium present, and therefore it appeared to be a property of the uranium atom itself. Mme Curie carried out an elaborate investigation of the phenomenon, and found that the uranium ore used was much more active, in proportion to the amount of uranium present, than a pure uranium compound, and accordingly inferred that some other substance, still more active, was present. She succeeded in 1898 in isolating this substance, which proved to be a new element, and was given the name of Radium.

5·301. An astonishing fact was soon discovered about the occurrence of radium. It occurs in nature only in the presence of uranium, which itself never occurs without radium. The ratio of the masses of the two elements present in a sample of ore is almost always the same, except perhaps in some of the most recent rocks, namely $(3{\cdot}40 \pm 0{\cdot}03) \times 10^{-7}$ parts of radium to one part of uranium*. Such a constancy suggests a chemical combination, but the atomic weights of uranium and radium are respectively about 238 and 226, and therefore one atom of radium would

* Rutherford and Boltwood, *Amer. J. Sci.* **22**, 1906, 1–3; S. C. Lind and L. D. Roberts, *J. Amer. Chem. Soc.* **42**, 1920, 1170.

have to unite with about three million atoms of uranium to give the proper ratio. Such a complexity is not approached by the most complicated chemical compounds known, so that the chemical hypothesis is most unplausible.

5·302. A further discovery led to the explanation. Radium itself was found to undergo a gradual change. A mass of a radium compound enclosed in a sealed vessel was found to liberate a gas called 'radium emanation,' the rate of formation of the gas being simply proportional to the amount of radium present. The rate was such that, if initially one gram of radium was present, only half a gram would be present 1500 years afterwards. The rest would be transformed into the emanation and into the further disintegration products of the emanation. All uraniferous ores are many thousands of years old, on any geological hypothesis, and therefore we have to explain how it is that any radium exists at all: why it has not all broken up long ago. The explanation suggested by its invariable association with uranium is that as fast as it breaks up new radium is formed by the break-up of the uranium itself. The suggestion was experimentally verified by Soddy*, who prepared a specimen of uranium quite free from radium, kept it for some years, and was able to demonstrate the presence of radium in the specimen at the end of the experiment.

5·31. Uranium, however, does not pass straight to radium, nor is the emanation the final product. The latter, in fact, survives only a few days. Suppose then that u atoms of uranium are present at time t, and suppose that each atom of uranium becomes in succession unit amounts of various recognizable stages $X_1, X_2, X_3, \dots X_n$. Suppose the numbers of uranium atoms that have gone to form the amounts of these products present at the instant considered to be $x_1, x_2, x_3, \dots x_n$. Further suppose that what has been proved to be true of radium is true in general, namely that the rate of break-up of any product is simply proportional to the quantity present†, and accordingly that any product X_r generates in unit time $\kappa_r x_r$ units of the next product X_{r+1}. The rate of degeneration of atoms of uranium itself will similarly be denoted by κu. Then $u, x_1, x_2, \dots x_n$ satisfy the following differential equations:

$$\frac{du}{dt} = -\kappa u \qquad \text{.........(1)},$$

$$\frac{dx_1}{dt} = \kappa u - \kappa_1 x_1 \qquad \text{.........(2)},$$

$$\frac{dx_2}{dt} = \kappa_1 x_1 - \kappa_2 x_2 \qquad \text{.........(3)},$$

....

$$\frac{dx_n}{dt} = \kappa_{n-1} x_{n-1} \qquad \text{.........(4)}.$$

* *Phil. Mag.* **9**, 1905, 768–779; **16**, 1908, 632–638; **18**, 1909, 846–865; **20**, 1910, 340–349.

† Rutherford and Soddy, *Phil. Mag.* **5**, 1903, 576–591.

Suppose that initially there are no degradation products present, so that when t is zero

$$u = u_0;\ x_1 = x_2 = x_3 = \ldots = x_n = 0 \qquad \ldots\ldots\ldots(5).$$

The solutions of these equations are

$$u = u_0 e^{-\kappa t} \qquad \ldots\ldots\ldots(6),$$

$$x_1 = \frac{\kappa u_0}{\kappa_1 - \kappa}(e^{-\kappa t} - e^{-\kappa_1 t}) \qquad \ldots\ldots\ldots(7),$$

$$x_2 = \frac{\kappa\kappa_1 u_0}{\kappa_1 - \kappa}\left\{\frac{1}{\kappa_2 - \kappa}(e^{-\kappa t} - e^{-\kappa_2 t}) - \frac{1}{\kappa_2 - \kappa_1}(e^{-\kappa_1 t} - e^{-\kappa_2 t})\right\} \ldots\ldots(8),$$

and so on, the expressions becoming more and more complicated as later products are considered*. If, however, all the degeneration products are short-lived in comparison with uranium, so that $\kappa_1, \kappa_2, \ldots \kappa_{n-1}$ are all great in comparison with κ, and t is so great that $1/t$ is less than the smallest of $\kappa_1, \kappa_2, \ldots \kappa_{n-1}$, the solutions reduce approximately to

$$u = u_0 e^{-\kappa t} \qquad \ldots\ldots\ldots(9),$$

$$x_1 = \frac{\kappa u}{\kappa_1} \qquad \ldots\ldots\ldots(10),$$

$$x_2 = \frac{\kappa u}{\kappa_2} \qquad \ldots\ldots\ldots(11),$$

$$\ldots\ldots\ldots\ldots\ldots\ldots$$

$$x_{n-1} = \frac{\kappa u}{\kappa_{n-1}} \qquad \ldots\ldots\ldots(12),$$

$$x_n = \int_0^t \kappa_{n-1} x_{n-1}\,dt = u_0(1 - e^{-\kappa t}) \qquad \ldots\ldots\ldots(13).$$

Thus the amounts of all products present except the last remain in fixed ratios to each other and to the amount of uranium left, the ratios being such that the number of units of any product that break up in a given time is the same for all products and equal to the number of atoms of uranium that break up in that time. Thus we have an explanation of the constancy of the radium/uranium ratio. Further, the value of this ratio enables us to find κ. We know from experiments on the rate of disintegration of radium that every year 1/2280† of the radium present breaks up. If r is the number of radium atoms present in a rock specimen, and we allow for the difference in atomic weight between uranium and radium, we find (observing as is natural that each atom of uranium ultimately gives one atom of radium) that

$$\frac{r}{u} = \frac{3{\cdot}4 \times 10^{-7} \div 226}{1 \div 238} = 3{\cdot}58 \times 10^{-7} \qquad \ldots\ldots\ldots(14),$$

whence

$$1/\kappa = 6{,}370{,}000{,}000 \text{ years} \qquad \ldots\ldots\ldots(15).$$

* The solution, and approximations to it, are most easily obtained by Heaviside's methods. Cf. Jeffreys, *Operational Methods in Mathematical Physics*, Camb. Univ. Press, 1927, 35–38.

† V. F. Hess and R. W. Lawson, *Wien. Sitzungsber.* 127, 1918, 1–55.

Knowing the rate of break-up of uranium, we shall now be able to find the time since the formation of any rock if we know the amounts of uranium and of the end product present. If indeed l denote the number of units of the end product present, we shall have

$$t = \frac{1}{\kappa} \log \frac{u + l}{u} \qquad \text{.........(16)},$$

and if l/u is small an approximation to this will be

$$t = \frac{l}{\kappa u} \qquad \text{.........(17)}.$$

Thus a chemical analysis of the rock should give its age when the end product is identified.

5·32. The argument so far given is independent of whether the various disintegration products are pure substances or not. All that has been assumed about their constitution is that each of them is made up of units of similar composition, each unit having been derived from one atom of uranium. The units themselves may be composed of atoms, which need not be all alike. Thus the occurrence of several chemically different substances in any disintegration product is possible. Experiments on the behaviour of uranium and radium have shown that this is actually the case. The gaseous emanation from radium contains two substances, namely the inert gas helium, which undergoes no further change, and a radioactive gas called radon (niton); each atom of radium yields one atom of each of these gases. Radon again breaks up, each atom giving one atom of helium and one of a further transient substance called Radium A. The disintegration continues and no fewer than five atoms of helium are lost in succession from a single atom of radium. Now the atomic weight of radium is 225·95. That of helium is 4·002*. Thus the fifth product of the disintegration of an atom of radium should be five atoms of helium and one atom of an element with an atomic weight of 205·94; or possibly the heavier product might be two similar or even dissimilar atoms. The direct identification of this substance, or these substances, by keeping a sample of radium until an analysable quantity of the end product has accumulated, has not yet been carried out, but indirect evidence has given very definite information about its nature. Before we proceed to this point, however, we notice that the atomic weight of uranium, according to the best determinations, is 238·14†. The difference between the atomic weights of uranium and radium is nearly three times the atomic weight of helium, and it is actually found that uranium gives off three helium atoms in succession, forming the elements Uranium II, Ionium, and Radium. Thus the unit of the third disintegration product of uranium is an atom of radium and three

* Aston, *Nature*, **120**, 1927, 958.

† O. Hönigschmid and W. E. Schilz, *Zs. f. anorg. Chem.* **170**, 1928, 145–160.

atoms of helium, and the unit of the eighth product is eight atoms of helium and one or more atoms the sum of whose atomic weights is nearly 206·0.

5·321. No element with an atomic weight differing from 206·0 by less than the probable error of an atomic weight determination was known when radioactivity was discovered. The nearest were Bismuth 208, Lead 207·1, and Thallium 204. If the product contained two similar atoms of atomic weight 103, the conditions would be satisfied. The elements whose atomic weights are nearest to this are Rhodium 102·9, and Ruthenium 101·7. The only one of these elements, and indeed the only metal at all other than the known disintegration products, invariably found in uranium minerals is lead. We have therefore strong reason for believing that the final product of the disintegration of uranium is lead. The discrepancy in atomic weight still presented a difficulty, until direct determinations of the atomic weight of lead from uranium minerals were made in 1914 by Hönigschmid and Fräulein St Horovitz, Richards and Lembert, and Maurice Curie. It was found to be 206·2, not far from the predicted atomic weight, and almost a whole unit lower than that of ordinary lead. Later determinations have given values near 206·05*. Thus the end product is identified; its unit consists of an atom of this new kind of lead, which will be called uranium lead, and eight atoms of helium. Like ordinary lead, uranium lead is not radioactive; no further degeneration occurs after this stage. Thus the determination of the age of a uranium mineral requires the determination of the amount of uranium still present, and of the amount of helium or of uranium lead present. When these are known, the ratio l/u is determinable, and then the age of the mineral can be found from 5·31 (17).

5·33. The use of the lead/uranium ratio for finding the ages of minerals was first attempted by Boltwood†, who found that the ratio of the amounts of uranium and lead present in uranium minerals of the same geological age was approximately constant. The helium/uranium ratio was applied in 1908–10 by the present Lord Rayleigh, then the Hon. R. J. Strutt. Both methods have been extensively used by Holmes. There is little doubt that the method involving the use of lead is the superior. It will be seen that the applicability of either method in any particular case depends on whether three conditions are satisfied. First, the final product estimated must have been absent from the mineral when this was first formed. There seems no reason to believe that original helium ever occurs in appreciable quantities in igneous rocks; original lead is common, but it is possible in many cases to attach a very high probability to its absence. Uranium in pitchblende is in the form of the oxide uranous uranate

* O. Hönigschmid and L. Birkenbach, *Ber. d. d. Chem. Ges.* **56**, 1923, 1837.

† *American Journal of Science*, **23**, 1907, 77–88.

$U(UO_4)_2$; no lead compound isomorphous with this occurs in these ores, and hence the lead and uranium must crystallize separately. When the crystals are too small for an analysis of a single crystal to be undertaken, as indeed is usually the case, it is more difficult to be sure that no crystals of a lead compound are intermingled with those of the uranium compound. If, however, we confine our attention to ores containing a large percentage of uranium, we can be practically certain that the amount of original lead is small compared with the amount of uranium. Doubt can in any case be dispelled or established by an atomic weight determination.

5·331. The second condition required is that radioactivity must be the only agency that has altered the composition of the mineral since it was formed. All the lead or helium generated in the mineral must still be in it. Thus minerals altered by heat or water must be excluded, since heat produces recrystallization and accordingly separation of lead from the associated uranium, and promotes the diffusion of helium through the rock or even into the free air, while water may produce a chemical separation. If the mineral becomes exposed to the air, loss of helium by diffusion into the air is certain, and even within the crust leakage into the surrounding rocks is probable. Thus a mineral that has not undergone metamorphosis by heat or water probably contains its proper amount of lead; but it is very doubtful whether any mineral contains the whole of the helium generated from its uranium. Thus estimates based on the helium/uranium ratio will be systematically lower than the true ages of the rocks.

5·332. Third, it must be possible to determine accurately the amount of lead or helium in the final product. The estimation of lead is not a difficult process, and presents no likelihood of serious error. In estimating helium, however, the mineral has to be ground to a fine powder, which then has to be heated *in vacuo* to drive off the included helium. Leakage occurs to some extent during the powdering process, and on this ground again the age found from the helium/uranium ratio must be too low.

We thus see that while, with proper caution in selecting the minerals to be examined, the lead/uranium ratio is likely to give correct determinations of the ages of rocks, the helium/uranium ratio is practically certain to give results systematically too low. Thus estimates made by the latter method can be regarded only as lower limits to the possible ages of the rocks they refer to.

5·34. Uranium and the elements derived from it are not the only radioactive substances known. The element thorium, of atomic weight 232·15, is radioactive, and produces helium in its degeneration just as uranium does, one atom of it liberating six atoms of helium in succession. The final product is a lead of atomic weight about 208. The identification of the final product is based on the fact that lead is the only heavy metal

other than thorium and uranium regularly found in thorium minerals. The calculated atomic weight is 208·14. Lead from minerals containing much thorium is found to have a higher atomic weight than ordinary lead, but unfortunately these minerals are seldom free from uranium, and the results obtained are therefore below the theoretical atomic weight. The purest thorium mineral analysed was found by O. Hönigschmid* to contain lead with an atomic weight of 207·9.

Thorium was found by Rutherford and H. Geiger† to break up at a rate of one part in $1{\cdot}87 \times 10^{10}$ per year. Consequently it might be thought that it could be used, like uranium, for the measurement of geological time. There are, however, rather serious difficulties.

5·35. As thorium is usually associated with uranium, it is best to have first a formula for calculating the age of a mixed mineral from its lead content. Equation 5·31 (15) implies, when we allow for the difference of atomic weights of uranium and lead, that $7{\cdot}37 \times 10^9$ grams of uranium should produce 1 gram of lead per year. Thus if a uranium mineral free from thorium contains x grams of uranium and z grams of lead, the age of the mineral is Cz/x, where C is $7{\cdot}37 \times 10^9$ years, provided z/x is a small number. Similarly $20{\cdot}8 \times 10^9$ grams of thorium produce one gram of lead per year, and therefore a mineral containing x grams of uranium and y grams of thorium is accumulating lead at the rate of

$$\frac{x}{7{\cdot}37 \times 10^9} + \frac{y}{20{\cdot}8 \times 10^9} = \frac{x + 0{\cdot}35y}{7{\cdot}37 \times 10^9} \qquad \text{.........(1)}$$

grams per year. Hence if z is the number of grams of lead in the mineral, the age should be $Cz/(x + 0{\cdot}35y)$. Holmes and Lawson‡, as the result of a discussion taking account of the effects of possible systematic errors, give the approximate formula

$$\frac{z}{x + 0{\cdot}38y} \times 7400 \text{ million years} \qquad \text{.........(2).}$$

A small correction is needed for very old minerals. If

$$v = 1{\cdot}155 \frac{z}{x + 0{\cdot}38y},$$

the estimate in (2) should be multiplied by $(1 - \tfrac{1}{2}v + \tfrac{1}{3}v^2)$.

When this formula is applied to minerals containing little thorium, it is ordinarily found to give results consistent with the ages determined by geological methods; of two minerals, that known to be the older from stratigraphical evidence has the greater lead/uranium ratio. Exceptions are few, and usually easily explicable as due to original lead or metamorphism. Minerals rich in thorium, however, give very irregular results, some being as great as those found from uranium minerals of the same age,

* *Zs. f. Electrochem.* **22**, 1916, 18.

† *Phil. Mag.* **20**, 1910, 691.

‡ *Amer. J. Sci.* **13**, 1927, 327–344.

but many very much lower (down to about a quarter). The variations are not systematic; two thorium minerals with the same age and associated with uranium minerals having the same Pb/U ratio may have very different ratios of the type $z/(x + 0{\cdot}38y)$. This fact seems, as Holmes points out*, to exclude the possibility that the variations are due to any peculiarity in the behaviour of uranium or thorium, such as Joly favours, and to show that they are due to differences in the history of the individual specimens. Lead generated in a uranium mineral would come into existence surrounded by atoms of uranium and oxygen, and would probably form instantly the highly insoluble lead uranate. On the other hand thorates do not exist, and thorium lead would form only an oxide or a silicate. Subsequent leaching by water would therefore tend to remove thorium lead and not uranium lead. Holmes, by a discussion of the atomic weights of specimens of lead from mixed minerals, is able to show that the ratio of the amounts of thorium lead and uranium lead in them is less than would be expected from the amounts of thorium and uranium, and to different extents strongly suggesting that thorium lead has been removed. Consequently results based on minerals containing much thorium are to be treated with caution.

5·351. The following table gives determinations of the ages of minerals over a wide range of geological time, taken from the above paper of Holmes and Lawson. The ages are given in millions of years.

Geological Age	Absolute age	Geological Age	Absolute age
Late Oligocene	37	Late pre-Cambrian (?)	587
Late Cretaceous (?)	59	Upper pre-Cambrian	640
Permo-Carboniferous	204	Middle pre-Cambrian	987–1087
Permian to Devonian	239–374	Lower pre-Cambrian	1257

The table shows many serious gaps. For the whole Mesozoic era there is no determination whose geological age is definitely known, and there are uncertainties about the geological age of most of the Palaeozoic specimens. The trouble is the difficulty of fixing accurately the geological age of an igneous rock. The oldest rock mentioned is from Australia. Others from Mozambique and South Dakota have given greater ages (the latter 1525 million years), but are considered less trustworthy by Holmes†. It must be noticed that these rocks are intrusive into older sediments, and the age of the earth itself must be greater than that of the oldest rock analysed.

5·36. The lead/uranium ratio forms the basis of a different method, due to Prof. H. N. Russell‡. If we assume that all the lead of average igneous rocks has been derived from uranium and thorium since the formation of the earth, we shall obtain an estimate of the age of the crust as a whole. But as there was probably some lead in the crust initially

* *Phil. Mag.* **1**, 1926, 1055–1074.
† Cf. also A. C. Lane, *Science*, **67**, 1928, 631.
‡ *Proc. Roy. Soc.* **99**, 1921, 84–86.

this estimate will be in excess of the truth. Holmes has revised Russell's work with better data*. The proportions of uranium, thorium, and lead in average igneous rocks are respectively 6×10^{-6}, 15×10^{-6}, and $7{\cdot}5 \times 10^{-6}$. If thorium was absent and all the lead uranium lead, the original amount of uranium must have been $14{\cdot}7 \times 10^{-6}$, and the age of the crust therefore 5700 million years. Holmes, allowing for thorium, gets 3200 million years. Aston, again, has recently succeeded in separating ordinary lead (from ores) into its isotopes, and has shown that it consists of three components, of atomic weights 206, 207, and 208†, in the proportions 4 : 3 : 7. The first is uranium lead, and the last thorium lead. The intermediate one cannot be of radioactive origin, and if it is present to any important extent in the lead generally distributed through rocks a further reduction in the estimated age will be necessary. It appears that the lower limit to the age of the earth obtained by studying individual minerals, and the upper limit given by the constitution of the crust as a whole, are sufficiently close together to fix the age as roughly 2000 million years.

5·4. *The Denudational Methods.* We come now to the second and third methods of estimating geological time, which are usually described together as the geological methods; but since this may be held to constitute too narrow a use of the word 'geological,' they will be called the 'denudational' methods in the present work. Their methodological footing is altogether inferior to that of the method based on radioactivity. It has already been explained that an estimate of geological time requires two conditions: we must know the law satisfied by the rate of change we are using as our standard, and we must know its total extent in the interval we are measuring. The former condition is fulfilled by no denudational method; the latter is probably fulfilled with moderate accuracy by the accumulation of salt in the sea, but certainly not by the formation of sediments. Considering first the law of the change, we know the present rates of transport of salt and detritus to the sea by rivers, but it is not known how these rates have varied in the past. The rate of land erosion must depend on the slope of the land, the quantity, temperature, and carbon dioxide content of the rain falling, and on the nature of the soil exposed. No quantitative relation is known between the rate of denudation and any one of these factors, nor do we know even approximately how any one of these factors themselves has varied during geological time. We have a good deal of information relating to the type of rocks exposed in many places at various geological dates, but in most places the nature of the solid surfaces has undergone considerable changes during geological time, and there is no geological date such that the nature of the solid surface then is known for all points of the earth. We often know whether the land in some region was rising or sinking at a particular date, but we never know the precise

* *Nature*, 117, 1926, 482. † *Ibid.* 120, 1927, 224.

extent of the elevation nor of the change of slope. Information concerning the amount and nature of the rainfall is still more vague in character. Finally, even if we had all this information, we should still not be able to find the rate of denudation at any geological date, since the physical laws connecting it with the relevant data are still unknown.

5·41. In the estimates hitherto made by the denudational methods*, it has always been assumed that the rate of denudation has been uniform throughout geological time, an assumption that is incorrect for the reasons just given. The amount of sodium carried to the sea annually is about $1{\cdot}56 \times 10^{14}$ grams, and the total mass of the sodium in the ocean is about $1{\cdot}26 \times 10^{22}$ grams. If the accumulation had been uniform, the age of the ocean would have been 8×10^7 years. This is practically Joly's estimate of 1899. But much of the sodium carried to the sea is derived from the denudation of sedimentary rocks, and has therefore been in the sea before. Igneous rocks contain only about 2 per cent. of the chlorine required to combine with their sodium, and therefore it is probable that nearly all the chlorine in the ocean is of volcanic origin. If so, the amount of sodium corresponding to the amount of chlorine carried to the sea by rivers must be almost wholly derived from sedimentary rocks. This amounts to about 60 per cent. of the whole amount of sodium carried by rivers. Hence the amount of new sodium is unlikely to exceed $6{\cdot}9 \times 10^{13}$ grams annually. If this value is adopted, the corresponding age of the ocean is $1{\cdot}8 \times 10^8$ years. This is still too low, for much unchlorinated sodium must also be included in the sedimentary rocks, so that some of even the unchlorinated sodium must have been in the ocean before. Hence a further increase by a practically incalculable amount is necessary.

Holmes points out† further that the current estimates of the sodium carried by rivers in solution alone are greater than the whole sodium content of the rocks being denuded. The explanation seems to be that the sodium is usually determined by difference and not directly, and is therefore affected by the systematic errors of some of the other metals. Consequently the method is vitiated from the start.

5·42. The method based on the accumulation of sediments also meets with additional difficulties. After an elaborate discussion of the possible ways of utilizing it, Holmes decides that the most satisfactory is probably as follows. The igneous rocks at present exposed at the earth's surface produce altogether a cubic mile of sediments in five years. The total volume of sediments on the earth's surface is estimated at 7×10^7 cubic

* Most of the following arguments are from *The Age of the Earth*, by A. Holmes, Harpers, 1913. Full references to earlier work on these lines will be found there.

† *Phil. Mag.* **1**, 1926.

miles. Sediments derived from other sediments are not new, and are excluded by this method. The age of the ocean is thus estimated at $3{\cdot}5 \times 10^8$ years.

5·43. These two methods are, however, incapable of giving satisfactory determinations of the age of the ocean, for the reasons already given. The results are much smaller than the age of the oldest known rocks whose lead/uranium ratios have been determined, the difference being too great to be explicable by uncertainty as to the present rate of denudation or the actual total extent of denudation. Accordingly they are to be regarded as in error owing to variations in the rate of denudation. Their value, such as it is, is that they amount to a proof that the present rate of denudation is several times greater than the average of the past; they are not estimates of the age of the ocean.

5·5. *The Solar Energy Method.* The fourth method is the original method of Lord Kelvin*, based on the contraction theory of Helmholtz†. If m be the mass of a body and U the gravitational potential at its surface, the kinetic energy acquired by a mass dm in falling from an infinite distance to the surface of the body is $U dm$. When the added mass reaches the surface the kinetic energy becomes converted into heat and hence becomes available for radiation. Thus the total energy liberated by condensation can be found by supposing the whole mass to be brought up gradually from an infinite distance and deposited in thin uniform layers over the surface, and adding up the energies acquired by all in their fall. If the mass be supposed of uniform density ρ in its final stage, and if the radius at any intermediate stage be r, and the final radius a, we find for the energy of condensation the amount

$$\begin{aligned} W &= \int \frac{fm}{r}\, dm = \int_0^a \tfrac{4}{3}\pi f \rho r^2 \,.\, 4\pi\rho r^2 dr \\ &= \tfrac{16}{15}\pi^2 f \rho^2 a^5 \\ &= \frac{3}{5}\frac{fM^2}{a}, \end{aligned}$$

where M is the final mass. In the case of the sun this energy amounts to $2{\cdot}6 \times 10^{48}$ ergs, or $1{\cdot}3 \times 10^{15}$ ergs for each gram of the sun's mass. The latter result shows that any chemical energy in the sun must be of very small importance in comparison with condensational energy, since the energy of the most violent chemical reactions amounts only to quantities of the order of 10^4 calories per gram, or of 10^{11} ergs per gram.

Now the radiation received by a square centimetre of material exposed normally to the sun's radiation at the earth's distance from the sun is 0·03 cal. per second. Taking the earth's distance from the sun as $1{\cdot}5 \times 10^{13}$ cm., we see that the sun must be losing energy at a rate of

* *Phil. Mag.* **11**, 1856, 516–518. † *Phil. Mag.* **23**, 1862, 158–160.

$3 \cdot 3 \times 10^{33}$ ergs per second. Thus the total condensational energy of the sun would provide for radiation at the present rate for $7 \cdot 8 \times 10^{14}$ seconds or $2 \cdot 5 \times 10^{7}$ years.

5·6. *Possible Sources of Solar Energy.* This estimate is only of the order of a fiftieth of the age of the oldest rocks, as found from the lead/uranium ratio. Accordingly either some other, and much more abundant, supply of energy is available in the sun, or else there is a definite inconsistency between two results both based on physical laws. The geophysical evidence was the first to indicate that there must be some error in the Kelvin scale of time in stellar evolution, but several purely astronomical arguments have been found to point in the same direction. Individually perhaps none of them is quite decisive, but together they provide a means of coordinating so many facts that they must be taken very seriously. The first was based on the period of the variable star δ Cephei. It has been seen that, given the mass and radius of a star, we can find its condensational energy (apart from a correction for heterogeneity, which does not affect the order of magnitude). If in addition we know the rate of emission of energy (that is, the bolometric absolute luminosity) we can find how fast the star should be contracting if its whole output of energy is derived from condensation. Now the star in question executes a regular pulsation, whose period is theoretically proportional to the inverse square root of the density. Thus condensation should alter the period at a calculable rate. According to Prof. Eddington's latest estimate* the decrease should be 17 seconds annually. Estimates of the observed change differ, but none exceeds 0·1 second annually. Accordingly the evolution cannot be proceeding more than $\frac{1}{170}$ as fast as the contraction theory indicates. Other Cepheid variables point to the same conclusion†.

What brought the matter to a head was Eddington's theory of stellar luminosity, published in 1924‡. It was found both by theory and observation that the luminosity of stars is almost wholly determined by their mass and to a very subsidiary extent by anything else. The stars capable of being tested ranged in absolute magnitude from -4 to $+11$, and none deviated from the luminosity calculated from the mass alone by much over one magnitude. Now on the theory of relativity all forms of energy possess mass; when a star radiates away a quantity of energy W it loses mass W/c^2, where c is the velocity of light. Important loss of mass from a star, so far as we know, can take place in no other way§. Hence stars are losing

* *The Internal Constitution of the Stars*, 1926, 290.

† The pulsation theory of Cepheid variation is due to Prof. Eddington, *M.N.R.A.S.* **79**, 1918, 2–22, and Prof. H. Shapley, *Ap. J.* **40**, 1914, and *Mount Wilson Contributions*, Nos. 92, 153, 154. It is not free from difficulties, and is definitely rejected by Jeans. Cf. *M.N.R.A.S.* **87**, 1927, 326–330 for a summary of the arguments on both sides.

‡ *M.N.R.A.S.* **84**, 1924, 308–332.

§ Direct loss by expulsion of atoms was discussed by E. A. Milne, *M.N.R.A.S.* **86**, 1926 459–473.

mass at rates determined by their luminosities, and therefore by their masses. This gives at once a soluble differential equation to find out how the mass and luminosity of a star vary with the time. Eddington* finds the following durations, among others, of stages in the evolution of a star:

Absolute bolometric magnitude	Mass (Sun = 1)	Duration (10^{10} years)
0·0 to + 2·5	3·7 to 1·73	93
+ 2·5 to + 5·0	1·73 to 0·92	521
+ 5·0 to + 7·5	0·92 to 0·53	3630

Thus we are at once forced to think in terms of 10^{12} or 10^{13} years as our unit of time instead of 10^7 years. The sun is losing energy at a rate of $3{\cdot}3 \times 10^{33}$ ergs per second, and therefore mass at $3{\cdot}7 \times 10^{12}$ grams per second. The loss of mass in 10^{10} years would be about 10^{30} grams, or $\frac{1}{2000}$ of the mass of the sun. It is for this reason that no appreciable change in the state of the sun during geological time is to be expected†.

The argument assumes that the evolution of stars is along a series such that Eddington's mass-luminosity relation holds at all points. The alternative would be that each star satisfies the relation for only a short time and is too faint to be observed before and after the interval when it satisfies it. There is, however, no direct evidence for such a view, and even if it were correct it would still vitiate the simple theory of condensation.

The extended time-scale also affords hope of accounting for the large eccentricities of the orbits of binary stars and for the correlation between the masses of stars and their velocities. Both phenomena could be accounted for by frequent close approaches to other stars; the new time-scale seems to provide time for enough such approaches‡, while shorter scales did not.

The most probable mechanism of the loss of mass seems to be that in stellar conditions a positive nucleus (hydrogen ion) and a negative one (electron) may coalesce. Their masses arose from their electric charges; but these neutralize each other and go out of existence, nothing being left but an electromagnetic wave travelling out with the velocity of light. This idea was offered originally by Jeans§ as an explanation of radioactivity. It is of course known now that this involves no losses of mass of the order of a unit in atomic weight, but the phenomenon of explaining something different from what one sets out to explain is well known in theoretical physics‖. It was tentatively revived by Eddington in 1917¶, who also considered a less drastic hypothesis. The atomic weight of hydrogen, on the basis O = 16, is 1·008. All other atomic weights of single isotopes are nearly integers. If then other elements are composed of

* *Internal Constitution of the Stars*, 309.

† Cf. *M.N.R.A.S.* 85, 1925, 413–416.

‡ Jeans, *M.N.R.A.S.* 85, 1924, 2–11.

§ *Nature*, June 2, 1904, 111.

‖ The losses of energy in radioactive changes correspond to differences of atomic weight of the same order of magnitude as the actual departures of the atomic weights from whole numbers, and must be to some extent responsible for them.

¶ *Nature*, Aug. 2, 1917, 445.

hydrogen nuclei and electrons, there has been a loss of mass of 0·8 per cent., which has presumably gone as energy*. This would supply the sun's expenditure of energy for $1{\cdot}4 \times 10^{11}$ years, enough for geophysical purposes, but it does not explain the relevant astronomical facts. Accordingly the conversion of matter into energy seems to be the most satisfactory hypothesis yet offered, though it cannot yet be said to be fully understood.

The method based on tidal friction will be described in Chapter XIV; it gives the order of magnitude of the age of the moon. The thermal method is vitiated by the effects of radioactive heating (see Chapter VIII); it can however be made to yield a lower limit to the age of the earth†, of the same order as that given by 5·5.

5·7. Thus of all the methods suggested for the measurement of the age of the earth, only the lead/uranium ratio method is quantitatively satisfactory. It implies that the age of the ocean exceeds about 1200 million years, and probably 1500 million years, but not 3000 million years. It is now necessary to discuss the lengths of the various stages in the history of the earth that elapsed before the formation of the ocean.

5·8. The time taken by the earth in liquefying and solidifying is easily shown to be a small fraction of its age. The rate of loss of heat per gram of its mass is given by the formula of 2·5, namely $3\sigma V^4/\rho a$, where a is the radius. If we take

$$\begin{aligned} \sigma &= 5 \times 10^{-5} \text{ c.g.s. Centigrade units,} \\ V &= 3000^\circ, \\ a &= 6{\cdot}4 \times 10^8 \text{ cm.,} \end{aligned}$$

so that we are taking practically the boiling point of terrestrial materials and the present radius, this rate of loss amounts to 4 ergs per second per gram, or 3 calories per gram per year. The total energy due to the initial temperature of the gaseous earth could hardly exceed 6000 calories per gram, while the formula in 5·5 for the energy due to condensation shows that the energy per unit mass is proportional to M/a, and therefore in the case of the earth must have been about 10,000 calories per gram. Thus if the earth remained gaseous the whole of the condensational and initial thermal energy would have been radiated away in 5000 years at most. The actual rate of radiation would be somewhat greater than this, since the earth in the gaseous state must have been more distended and therefore have had a larger radiating surface than at present. Thus liquefaction must have been complete within 5000 years of the formation of the earth.

If we suppose the primitive earth to have had within itself a store of energy per unit mass comparable with either of those suggested for the sun in 5·6, this energy would maintain radiation at the rate appropriate to a gaseous earth for 1/33 of the time it enabled the sun to radiate at the

* *Brit. Ass. Report*, 1920, 45.

† *Nature*, 108, 1921, 284.

same temperature; but the present rate of production of heat from atomic changes in the earth suggests that the atomic energy per unit mass in the earth is in no way comparable with what appears to be required for the sun.

5·81. The liquefaction would be complete when all the latent heat of evaporation had been radiated away, and cooling to the melting point would then proceed. The latent heat of fusion would then have to be lost before solidification was complete. The total loss of heat from the commencement of condensation to the completion of solidification would hardly exceed 2000 calories per gram. Taking the melting point as 1500°, we find that the rate of loss of heat per gram at that temperature would be $\frac{1}{5}$ calorie per year. Thus the time between the onset of solidification and its advancement to such a stage that internal convection was stopped* could not exceed 10,000 years. Hence we can allow in all 15,000 years from the formation of the earth to its solidification to this extent. The present temperature of the surface is almost wholly maintained by solar radiation, which has hardly changed, and it is therefore probable that the surface cooled to nearly its present temperature very soon after solidification. The formation of an ocean then became possible.

The formation of the moon can be placed with some accuracy with regard to the process just outlined. It has been seen that conditions suitable for the formation of the moon by resonance did not exist till the diameter of the earth had almost attained its present value: until that time rotation would be too slow for resonance. Hence it could not have taken place until liquefaction was nearly complete, at the earliest. The conditions could not occur after the solidification had proceeded far enough to stop internal fluid motions, for such solidification would introduce a good deal of rigidity (see p. 181), which is absent from a fluid, and thereby would considerably shorten the period of a free vibration. Thus the free vibration would again become quicker than the period of the semidiurnal tide; in other words, rotation would again be too slow for resonance. Hence we can say that the birth of the moon took place while the earth was almost wholly liquid, the amounts of gaseous and solid constituents present being insufficient to produce any considerable effect on its period of free vibration. The moon existed before the ocean, and was probably formed within 10,000 years of the formation of the earth as a separate body.

5·9. *Summary.* The earth probably became solid within 15,000 years of its ejection from the sun, and soon afterwards cooled at the surface sufficiently for an ocean to condense. The moon, if it was ever part of the earth, was formed about the time when solidification was starting. The

* A slight modification of this will become apparent in Chapter VIII, but need not affect the present discussion.

interval from the formation of the ocean to the present time constitutes the greater part of the age of the earth, and it is possible that all the previous stages together occupied only an insignificant fraction of it.

Collecting our results, we have found the following:

1. From the eccentricity of the orbit of Mercury, we saw that the whole time since the rupture is probably between 10^9 and 10^{10} years.

2. The ages of the oldest known minerals, found from the lead/uranium ratio, are about $1{\cdot}3 \times 10^9$ years, and since the geological evidence indicates that some sedimentary rocks are still older, the age of the ocean must exceed $1{\cdot}3 \times 10^9$ years.

3. The amounts of uranium, thorium, and lead in the crust as a whole indicate a time less than 3×10^9 years since solidification.

The second and third results are in good agreement with the first. The age of the earth is therefore probably between $1{\cdot}3 \times 10^9$ and 3×10^9 years. The other methods are less satisfactory. The best inference that can be drawn from the denudational methods is that the present rate of denudation is about four times the average of the past, a conclusion in harmony with the fact that the present time is just after a glacial period and a period of mountain formation, both of which would tend to increase the rate of denudation very considerably. The Kelvin method, based on the sun's supply of energy, leads to inferences about the stars that do not accord with the facts, and it therefore cannot be trusted. It indicates rather that the sun has some source of energy other than condensation; and it is possible that the mutual annihilation of positive and negative ions supplies most of this energy.

CHAPTER VI

Seismology : General Considerations and Structure of the Upper Layers

"I heard the water lapping on the crag
And the long ripple washing in the reeds."
TENNYSON, *The Passing of Arthur.*

6·1. *Petrological Preliminaries.* The rocks exposed at the surface of the earth fall into two main divisions, the igneous and sedimentary rocks. The latter are derived from the former, which will therefore be considered first. Igneous rocks are composed mainly of silicates and free silica. The silicates themselves are of three main types: orthosilicates, containing the quadrivalent radical SiO_4; metasilicates, containing the divalent radical SiO_3; and trisilicates, containing the quadrivalent radical Si_3O_8. Comparing them, we see that four metallic valencies correspond in an orthosilicate to one silicon atom, in a metasilicate to two, and in a trisilicate to three. This order is therefore one of increasing acidity, and can be continued by the addition of free silica, the acidic oxide, at the end. Some of the principal rock-forming minerals and their properties are given in the following table, which is mostly extracted from F. W. Clarke's *Data of Geochemistry*, 1924; G. W. Tyrrell's *Principles of Petrology*, 1926; Landolt-Börnstein's *Physikalische Tabelle*, 1923; and Boeke-Eitel, *Physikalisch-chemischen Petrographie*, 1923. Only a brief summary of the vast literature of petrology can of course be attempted here, but the solution of the problems treated in the present work has not yet proceeded so far as to make greater detail necessary.

	Name	Composition	Density	Melting-point (° C.)
Silica	Quartz	SiO_2	2·65	—
	Tridymite	SiO_2	2·3	1670
	Cristobalite	SiO_2	2·348	1710
Trisilicates	Orthoclase	$KAlSi_3O_8$	2·56	1200 (?)
	Albite	$NaAlSi_3O_8$	2·605	1100
Metasilicates	Diopside	$CaMg(SiO_3)_2$	3·275	1391
	Enstatite	$MgSiO_3$	3·1	1557 (?)
	Hypersthene	$(Fe, Mg)SiO_3$	3·4–3·5	1500–1550 (?)
Orthosilicates	Anorthite	$CaAl_2(SiO_4)_2$	2·765	1550
	Forsterite	Mg_2SiO_4	3·2	1890
	Fayalite	Fe_2SiO_4	4–4·14	1055–1075
	Muscovite	$KH_2Al_3(SiO_4)_3$	2·85	—
	Biotite	$KHMg_2Al_2(SiO_4)_3$	2·7	—
	Garnets	$(Ca, Mg, Fe)_3(Al, Fe)_2(SiO_4)_3$	3·4–4·2	—

With reference to this table, quartz is the commonest natural form of silica, but is transformed into tridymite at temperatures over 800° and

into cristobalite at 1200° (Day and Shepherd). Albite and anorthite form the isomorphous mixtures known as the plagioclase felspars. The common mineral augite is diopside, with part of the magnesium replaced by ferrous iron, and with some alumina (Al_2O_3) in solution. Enstatite, hypersthene and augite are known as pyroxenes; all are somewhat variable in composition and properties on account of the tendency to form isomorphous mixtures. Forsterite and fayalite mixed form the important mineral olivine. Natural specimens of this usually melt near 1400°. Muscovite and biotite are the two commonest micas. Hornblende is a mixture of the form $m\,Ca(Mg, Fe)_3(SiO_3)_4 + n\,CaMg_2Al_2(SiO_4)_3$. Several of the above minerals decompose at high temperatures, notably orthoclase, enstatite, muscovite, biotite, and the garnets, and the melting-points given are really temperatures of decomposition. The value 1557° given for enstatite corresponds to its isomer clinoenstatite; this partly decomposes on melting.

6·11. The principal rocks we shall need to consider are granite, granodiorite, syenite, diorite, basalt, eclogite, peridotite, and dunite. A typical granite consists mainly of orthoclase and free quartz, with small amounts of micas. Syenite is roughly similar, but does not contain free quartz. Diorite is essentially plagioclase felspar with some biotite, hornblende, or augite. In a granodiorite comparable amounts of quartz, orthoclase, and plagioclase are associated.

Basalt, dolerite, and gabbro are rocks of similar composition, the finest-grained ones being called basalts and the coarsest gabbros. The principal minerals in them are plagioclase, the pyroxenes, and often olivine and magnetite (Fe_3O_4). Peridotites are largely composed of olivine, dunite almost wholly so.

Eclogite consists largely of garnets. It is chemically equivalent to a gabbro, according to equations of the type

$$\underset{\text{anorthite}}{CaAl_2(SiO_4)_2} + \underset{\text{forsterite}}{Mg_2SiO_4} = \underset{\text{garnet}}{CaMg_2Al_2(SiO_4)_3}$$

$$\underset{\text{anorthite}}{CaAl_2(SiO_4)_2} + \underset{\text{diopside}}{CaMg(SiO_3)_2} = \underset{\text{garnet}}{Ca_2MgAl_2(SiO_4)_3} + \underset{\text{silica}}{SiO_2}$$

$$\underset{\text{albite}}{NaAlSi_3O_8} = \underset{\text{jadcite}}{NaAl(SiO_3)_2} + \underset{\text{silica}}{SiO_2}$$

Its interest arises from its high density; garnet has a density of about 3·7 and jadeite 3·34. Consequently eclogite is widely believed to be the normal form of gabbro at high pressures*.

Rocks are not constant in chemical composition. Every intermediate stage exists between a granite and a dunite. The names given to them represent a classification that geologists have found convenient, but two rocks in the same class may differ appreciably in composition.

The densities of typical rocks of the above types are: granite, 2·64;

* L. L. Fermor, *Records Geol. Survey of India*, **43**, 1913, 41–47.

granodiorite, 2·73; diorite, 2·85; syenite, 2·78; gabbro, 2·94; peridotite, 3·18; dunite, 3·3; eclogite, 3·4. The general correlation of density with basicity is to be specially noticed.

In addition many rocks exist in wholly or partly vitreous or glassy forms. Thus granite is chemically equivalent to obsidian, of density 2·33–2·41, and basalt to tachylyte, of density about 2·85.

6·12. Igneous rocks exposed at the surface are slowly acted upon chemically by water, aided by dissolved carbon dioxide. In a granite, for instance, the orthoclase is converted into kaolin and muscovite, thus:

$$\underset{\text{orthoclase}}{2KAlSi_3O_8} + 2H_2O + CO_2 = \underset{\text{kaolin}}{2HAlSiO_4.H_2O} + 4SiO_2 + K_2CO_3$$

$$3KAlSi_3O_8 + H_2O + CO_2 = \underset{\text{muscovite}}{KH_2Al_3(SiO_4)_3} + 6SiO_2 + K_2CO_3$$

The potassium carbonate and silica are removed in solution, and the kaolin and muscovite in suspension. Kaolin when redeposited produces clay. The disintegration of the quartz, mica, and unchanged orthoclase left is mainly mechanical, and gives ultimately sand with micaceous flakes. Similarly plagioclase is broken up, the sodium and calcium being removed in solution as carbonates or bicarbonates. Magnesium in rocks is also usually dissolved during denudation. Iron may be dissolved, but if so is usually oxidized and reprecipitated as ferric hydroxide, with the texture of a clay. Typical sandstones and shales are formed from the sands and clays produced in this way. The dissolved calcium and magnesium in the ocean are continually removed to form limestone and dolomite. The potassium, according to Clarke, is partly absorbed by clays and silts and partly precipitated on the ocean floor in glauconite*.

The densities of sedimentary rocks are naturally very variable. When freshly deposited they are loose and the pore space is filled with water; consequently the rocks are much less dense than the individual particles. When they are raised above sea level and drained the pore space is emptied without much contraction of the rock as a whole, and the density becomes still less. If a rock is metamorphosed by heat and pressure, pores close up; the density then increases, and in extreme cases becomes equal to that of a crystalline mineral of the same composition. The following data are taken mainly from Kempe's *Engineer's Year-Book* and the *American Mechanical Engineer's Pocket-Book*:

Clay, 1·8–1·93; slate, 2·77. Shales may have any intermediate density.

Sand, 1·39–1·9; sandstone, 2·2; quartzite, 2·6. (Cf. quartz, 2·65.)

Portland Stone (oolitic limestone), 2·37; limestone, 2·48–2·53; marble, 2·70. (Cf. calcite, 2·72.)

6·13. The commonest igneous rocks exposed at the surface are, in order, granite (with granodiorite), basalt, and andesite. The last is similar

* Cf. Holmes, *Geol. Mag.* 1919, 251–254, 340–350.

in composition to diorite, but is fine in texture. Now all these rocks have at some time or other come up from some depth within the crust (unless indeed some of the granite was originally at the top) and therefore are representative of the material at various depths. But we shall expect the materials to be more or less stratified according to density, and therefore, comparing these facts with the densities given in 6·11, we infer that the uppermost igneous layer in the crust is probably granitic, with a basalt layer below it, possibly separated by a layer of diorite. We shall see later, however, that this structure will not represent the whole of the facts revealed by seismology, and that especially a layer of some denser rock such as dunite is required below the basalt.

The idea of a general granitic layer, forming the primitive upper rocks of the continents, was given by Suess in *Das Antlitz der Erde*, 1885–1909. It depends on the widespread granitic areas that form the great continental shields, largely of pre-Cambrian ages, and on the abundance of sandstone, which must have been derived originally from granite. The denser rocks, such as basalt and andesite, have been intruded through the general granitic layer from greater depths. In addition large granitic intrusions of more recent age are known, formed presumably within the granitic layer itself and then driven up to the surface.

None of the rocks considered likely to be prevalent near the surface have densities over 3·4 or so. The mean density of the earth, on the other hand, is about 5·5. It can be shown that compression is unable to increase the mean density of the materials to anything like this amount, and a still denser material is indicated near the centre. Reasons will be given later for supposing this to be a core of nickel-iron alloy, in a liquid state, and with a density near 12. The density of such an alloy under ordinary pressures would be about 8; the difference is to be accounted for by the high pressure at great depths.

6·2. *Waves started by Fractures.* The earth's crust is continually undergoing permanent deformation under the influence of the stresses developed within it in the course of its evolution. The deformation may take the form of gradual flow or of fracture. In each case strain energy is converted during the set into energy of internal motion. In the case of flow, the transformation is gradual, and produces only local heating; however great its ultimate effect may be, it produces no immediately observable consequences. The behaviour of the crust in the event of fracture, on the other hand, may be illustrated by analogy with an elastic string stretched to such an extent that it ultimately snaps. The relief of tension at the point of rupture initiates a separation of the exposed ends, and the motion spreads through the whole length of the string. In the same way, a fracture in the crust of the earth starts a movement there, leading to a wave, which spreads throughout the earth.

6·3. *Elastic Waves within a Solid.* It is easily seen that two types of plane wave are possible in a homogeneous elastic solid body; by a plane wave we mean that the displacement of a particle at any instant depends only on the time and the distance from a fixed plane. If we take the axis of x perpendicular to this plane, and those of y and z in it, and if u, v, w are the components of displacement parallel to these axes, the equations of motion are

$$\rho \frac{\partial^2}{\partial t^2}(u, v, w) = (\lambda + \mu)\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right)\delta + \mu\nabla^2(u, v, w) \quad \ldots\ldots(1),$$

where ρ is the density, t the time, λ and μ the two elastic constants of the substance, and

$$\delta = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} \quad \ldots\ldots\ldots(2).$$

If we differentiate the three equations (1) with regard to x, y, z and add, we find

$$\rho \frac{\partial^2 \delta}{\partial t^2} = (\lambda + 2\mu)\nabla^2\delta \quad \ldots\ldots\ldots(3).$$

But since the wave, by hypothesis, is plane, all derivatives with regard to y and z are zero, and δ is simply $\partial u/\partial x$. Also the equation of motion in the x direction is

$$\rho \frac{\partial^2 u}{\partial t^2} = (\lambda + 2\mu)\frac{\partial^2 u}{\partial x^2} \quad \ldots\ldots\ldots(4),$$

and therefore if u satisfies (4), $\partial u/\partial x$ will satisfy (3). Further, if

$$\rho a^2 = \lambda + 2\mu \quad \ldots\ldots\ldots(5),$$

the general solution of (4) is

$$u = f(x - at) + g(x + at) \quad \ldots\ldots\ldots(6),$$

where f and g may be any functions whatever. If we imagine g to be zero, then if we increase t by an amount τ and x by $a\tau$, the value of u is unaltered. Thus the term in f represents a displacement such that after an interval τ every displacement is repeated at a distance $a\tau$ in the direction of x increasing: that is, it represents a wave travelling with velocity a in the direction of x increasing. Similarly the g term represents a wave travelling with the same velocity in the opposite direction.

The other displacements v and w satisfy the equations

$$\rho \frac{\partial^2}{\partial t^2}(v, w) = \mu \frac{\partial^2}{\partial x^2}(v, w) \quad \ldots\ldots\ldots(7).$$

It follows similarly that if $\qquad \rho\beta^2 = \mu \qquad \ldots\ldots\ldots(8)$

the displacements in v and w are represented by waves travelling parallel to the x axis with velocity β.

It appears therefore that if u is not constant, and δ therefore not zero, the disturbance will travel with velocity α. On the other hand if v and w are not zero the disturbance will travel with velocity β; and $\alpha > \beta$. The possible disturbances therefore break up into two types. One has the displacement in the direction of propagation, no displacement at right angles to this direction, and velocity α; the other has displacements only at right angles to the direction of propagation, and velocity β. For the latter $\delta = 0$. We can easily verify that in fact the systems of displacements

$$u = f(x - \alpha t), \; v = 0, \; w = 0 \qquad \text{.........(9)},$$

$$u = 0, \; v = \phi(x - \beta t), \; w = 0 \qquad \text{.........(10)},$$

$$u = 0, \; v = 0, \; w = \psi(x - \beta t) \qquad \text{.........(11)},$$

where f, ϕ, and ψ are any functions, are solutions of (1) and therefore represent possible types of waves. The type (9) are called the longitudinal, irrotational, condensational, primary, or P waves; (10) and (11) the transverse, distortional, equivoluminal, secondary, or S waves. Prof. H. H. Turner has very appropriately called them the 'push' and the 'shake.' The displacements v and w are independent; we can therefore have an S wave polarized in either direction in the wave-front. It is often useful to distinguish whether the displacement in an S wave is horizontal or in a vertical plane through the direction of propagation; in the former case it is called an SH, in the latter an SV wave. The two types are reflected differently at a horizontal interface between different materials.

The existence of the longitudinal and transverse elastic waves was first predicted theoretically by Poisson*. The case $\lambda = \mu$, making $\alpha = \sqrt{3}\beta$, is of theoretical interest, and is near the truth for much of the earth's material.

6·31. There is a general analogy between the properties of bodily elastic waves and those of sound or light. When they are not plane waves but spherical ones spreading from a symmetrical source in a uniform body, the displacement at a point as the wave passes decreases like the inverse distance from the source. In a heterogeneous body the propagation is in general approximately according to the laws of geometrical optics. If the origin is at a point F, the time of transmission of a P wave to a point Q on the surface by a given path would be $\int ds/\alpha$ taken along the path, where ds is the element of length along the path and α the local velocity of a P wave. The first P wave to arrive at Q is the one that travels by the path that makes this integral least. Similarly the first S wave to arrive comes by the path that makes $\int ds/\beta$ least. In a homogeneous body these paths are straight lines, but in a heterogeneous one they are convex towards the side where the velocity is greatest.

* *Mém. de l'Acad. Paris*, **8**, 1829, 623–627; **10**, 1831, 578–605.

6·32. We thus arrive at the notion of rays, or paths of least time, in seismology as in geometrical optics. The wave-fronts are the loci of points reached simultaneously by a wave of given type, and cut the rays at right angles. In dealing with the amplitudes of the waves we can often use an approximation based on the principle that energy is propagated without loss along the rays, so that if we consider a surface formed of rays, the energy that crosses every cross section of this surface is the same. This approximation is valid provided that the properties of the medium do not change appreciably within a single wave-length. That it is not exact can be seen by considering compressional waves of uniform amplitude propagated along the interior of a circular cylinder within a solid. The wave fronts are everywhere perpendicular to the axis, and therefore, on the above hypothesis, energy would be transmitted rigorously parallel to the axis, and no disturbance at all could be set up outside the cylinder. But as the wave travels along it sets up variable stresses on the surface of the cylinder, and these act on the matter outside the cylinder and set up motions in it. Thus a disturbance localized in a region will ordinarily be associated with a travelling disturbance in a fringing region. This phenomenon is *diffraction*. In general it is present whenever the wave-fronts have not spherical symmetry about a point; its effect on the distribution of energy is important in some seismological cases, but not in others. It does not in any case affect the time of transmission as determined from the minimal conditions.

6·33. Elastic waves undergo reflexion and refraction at interfaces between different media, and at free surfaces. The phenomena are, however, more complicated than in sound or optics, because each medium can in general transmit both longitudinal and transverse waves. Thus when a plane P wave is incident on a horizontal boundary between two solids it is ordinarily broken up into four waves, a reflected P and an SV wave, and a transmitted P and an SV wave. The two derived P waves will be referred to as 'of the original type' and the two SV waves as 'transformed waves.' Similarly an incident SV wave gives two SV waves and two P waves, the latter in this case being the transformed waves. An SH wave, on the other hand, gives only reflected and transmitted SH waves. Let us use the letter i to denote the angle made by the direction of travel of a P wave with the normal at the point of incidence, and j to denote that made by an S wave. Further let suffix 1 refer to the reflected waves and the original medium, and accents to the transmitted waves and the second medium. Then in the reflexion of a P wave the angles i and j are connected by the relations

$$\frac{\sin i}{\alpha} = \frac{\sin i_1}{\alpha} = \frac{\sin j_1}{\beta} = \frac{\sin i'}{\alpha'} = \frac{\sin j'}{\beta'} \quad \text{.........(1)},$$

where the waves referred to by the various expressions are, in order,

incident P, reflected P, reflected SV, transmitted P, transmitted SV. Similarly when the incident wave is of type SV we have the relations

$$\frac{\sin j}{\beta}=\frac{\sin i_1}{\alpha}=\frac{\sin j_1}{\beta}=\frac{\sin i'}{\alpha'}=\frac{\sin j'}{\beta'} \qquad \text{.........(2).}$$

If the incident wave is of type SH we have simply

$$\frac{\sin j}{\beta}=\frac{\sin j_1}{\beta}=\frac{\sin j'}{\beta'} \qquad \text{.........(3).}$$

It may happen that the sines of the angles i and j for some of the derived waves, as calculated from (1), (2), or (3), are greater than 1. The corresponding movement is then not a wave receding from the interface, but an oscillatory motion decreasing exponentially in amplitude as we recede from the interface. This case is analogous to total reflexion in optics. An ordinary derived wave carries energy continually away from the interface, the four together accounting for the whole of the energy brought by the incident wave; but these exponential disturbances transport no energy on an average.

Numerous cases of the reflexion and refraction of elastic waves have been worked out in detail*. The results are in many cases very remarkable. Thus for the reflexion of P waves at a free surface, when the ratio α/β is $\sqrt{3}$, at least half the energy goes into the reflected distortional wave for all angles of incidence between 27° and 88°. For normal and grazing incidence there is complete reflexion in the original type. When an SV wave is reflected, more than half the energy goes into the reflected P wave if the angle of incidence is between 15° and 35°, and the reflected distortional wave disappears altogether for incidence at 34° 16′ and 30°; but for incidence at any angle greater than 35° 16′ there is total reflexion in the original type. SH waves at a free surface are of course completely reflected in the original type.

At an interface between two media not too different in properties, the variation of the ratios of the amplitudes with angle of incidence is more regular. When a derived wave grazes the interface it always carries a vanishing fraction of the energy. Transmission in the original type is generally most complete at normal incidence. Transmitted transformed waves are usually small.

Where one of the media is a liquid, it transmits condensational waves only. A large condensational wave may be derived in this case from a distortional wave incident on a surface between a solid and a liquid.

6·34. In addition to the bodily waves so far considered, an elastic solid with a free surface can transmit waves such that the motion is

* C. G. Knott, *Phil. Mag.* (5) **48**, 1899, 64–97; E. Wiechert in *Erdbebenwellen I*, *Gött. Nach.* 1907, 415–549; B. Gutenberg in Sieberg's *Erdbebenkunde*, 1923; Jeffreys, *M.N.R.A.S. Geoph. Suppl.* **1**, 1926, 321–334. Knott's paper still covers the largest number of cases of geophysical interest.

greatest near the surface, and becomes inappreciable at a depth of a few wave-lengths.

If the origin be in the surface, the axis of z vertically downwards, that of x horizontal and in the direction of propagation, and that of y perpendicular to it, then in any harmonic wave propagated over the surface without change of type the component displacements must be the real parts of $(U, V, W)\, e^{\iota\kappa(x-ct)}$, where U, V, W are complex quantities which may be functions of z. Then substituting in 6·3 (3), which we derived without assuming that the wave was plane, we have

$$-\rho\kappa^2c^2\delta = (\lambda + 2\mu)\nabla^2\delta$$
$$= (\lambda + 2\mu)\left(\frac{\partial^2\delta}{\partial z^2} - \kappa^2\delta\right) \qquad \ldots\ldots\ldots(1).$$

The only solution of this equation that does not become infinite when z tends to infinity is

$$\delta = De^{-rz}\, e^{\iota\kappa(x-ct)} \qquad \ldots\ldots\ldots(2),$$

where D is a constant, and

$$r^2 = \kappa^2\left(1 - \frac{c^2}{\alpha^2}\right) \qquad \ldots\ldots\ldots(3).$$

Any displacement consistent with this value of δ must satisfy

$$-\rho\kappa^2c^2(U, V, W) = (\lambda + \mu)(\iota\kappa, 0, -r)De^{-rz} + \mu\left(\frac{d^2}{dz^2} - \kappa^2\right)(U, V, W) \quad \ldots(4),$$

since when R is of the form $Pe^{\iota\kappa(x-ct)}$, where P is a function of z alone,

$$\nabla^2 R = \left(\frac{d^2}{dz^2} - \kappa^2\right) Pe^{\iota\kappa(x-ct)} \qquad \ldots\ldots\ldots(5).$$

Also $\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z}$ is the real part of

$$\left(\iota\kappa U + \frac{dW}{dz}\right) e^{\iota\kappa(x-ct)} \qquad \ldots\ldots\ldots(6),$$

which accordingly must be equal to $De^{-rz}\, e^{\iota\kappa(x-ct)}$.

Particular solutions of these equations are found to be

$$(U, V, W) = -\frac{\alpha^2}{\kappa^2c^2}(\iota\kappa, 0, -r)\, De^{-rz} \qquad \ldots\ldots\ldots(7)$$
$$= (U_1, V_1, W_1), \text{ say} \qquad \ldots\ldots\ldots(8).$$

It is readily verified that

$$\iota\kappa U_1 + \frac{\partial W_1}{\partial z} = De^{-rz} \qquad \ldots\ldots\ldots(9).$$

Hence the complete solution, satisfying

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = \delta \qquad \ldots\ldots\ldots(10),$$

must be given by

$$(U, V, W) = (U_1, V_1, W_1) + (U_2, V_2, W_2) \quad \ldots\ldots\ldots(11),$$

where (U_2, V_2, W_2) satisfy the equations

$$-\rho\kappa^2 c^2 (U_2, V_2, W_2) = \mu\left(\frac{d^2}{dz^2} - \kappa^2\right)(U_2, V_2, W_2) \quad \text{......(12)},$$

and

$$\iota\kappa U_2 + \frac{dW_2}{dz} = 0 \quad \text{.........(13)}.$$

The solutions of these equations that tend to zero at an infinite depth are

$$(U_2, V_2, W_2) = (s, B, \iota\kappa)\, Qe^{-sz} \quad \text{.........(14)},$$

where Q and B are unspecified constants and

$$s^2 = \kappa^2 (1 - c^2/\beta^2) \quad \text{.........(15)}.$$

Now the boundary conditions are that there shall be no stress across the free surface; and if we neglect small quantities of the second order in the displacements this is equivalent to the condition that there shall be no stress over the plane $z = 0$. This gives

$$\frac{\partial u}{\partial z} + \frac{\partial w}{\partial x} = 0 \quad \text{.........(16)},$$

$$\frac{\partial v}{\partial z} + \frac{\partial w}{\partial y} = 0 \quad \text{.........(17)},$$

$$\lambda\delta + 2\mu\frac{\partial w}{\partial z} = 0 \quad \text{.........(18)}$$

when $z = 0$. Substituting in these equations from (8) and (14) we have

$$2\iota\frac{r}{\kappa}\frac{\alpha^2}{c^2} D = (\kappa^2 + s^2)\, Q \quad \text{.........(19)},$$

$$B = 0 \quad \text{.........(20)},$$

$$(\lambda + 2\mu)\left(1 - \frac{2\beta^2}{c^2}\right) D = 2\mu s\iota\kappa Q \quad \text{.........(21)}.$$

Eliminating D and Q, and cancelling a power of κ, we are left with an equation to find c, namely

$$\left(2 - \frac{c^2}{\beta^2}\right)^2 = 4\left(1 - \frac{c^2}{\alpha^2}\right)^{\frac{1}{2}}\left(1 - \frac{c^2}{\beta^2}\right)^{\frac{1}{2}} \quad \text{.........(22)}.$$

Rationalizing and cancelling a factor c^2 we have the cubic equation in c^2

$$\frac{c^6}{\beta^6} - 8\frac{c^4}{\beta^4} + 24\frac{c^2}{\beta^2} - 16\frac{c^2}{\alpha^2} - 16 + 16\frac{\beta^2}{\alpha^2} = 0 \quad \text{.........(23)}.$$

When c^2/β^2 is 0, the left side is $-16(1 - \beta^2/\alpha^2)$, which is always negative. When c^2/β^2 is 1, the left side is 1. Thus there is always a root between 0 and 1. Corresponding to this, real positive values of r and s exist, by (3) and (15). Since (23) has been obtained by rationalizing, the solution might be really a solution of (22) with the right side changed in sign; but when r and s are both positive this alternative is excluded, since both sides of (22) are positive. Thus there is always a possible type of surface waves with a velocity less than β.

When the solid is incompressible, α is infinite, and (23) has one real root,

$$c^2/\beta^2 = 0{\cdot}91275 \qquad \text{.........(24)},$$

whence
$$c = 0{\cdot}9554\beta;\ s = 0{\cdot}2954\kappa;\ r = \kappa \qquad \text{.........(25)},$$

$$u = A\,(e^{-\kappa z} - 0{\cdot}5433e^{-sz}) \sin \kappa\,(x - ct) \qquad \text{.........(26)},$$

$$v = 0,$$

$$w = A\,(e^{-\kappa z} - 1{\cdot}840e^{-sz}) \cos \kappa\,(x - ct) \qquad \text{.........(27)},$$

where A is a constant. The motion at the surface is given by

$$u = 0{\cdot}4567\,A \sin \kappa\,(x - ct) \qquad \text{.........(28)},$$

$$w = -\,0{\cdot}840\,A \cos \kappa\,(x - ct) \qquad \text{.........(29)},$$

so that the particles move in elliptic paths, the maximum vertical and horizontal displacements being in the ratio 1·9 : 1.

The two other roots are complex, namely

$$c^2/\beta^2 = 3{\cdot}5436 \pm 2{\cdot}2301\iota \qquad \text{.........(30)},$$

whence
$$r = \kappa;\ 4s/\kappa = (2 - c^2/\beta^2)^2$$

$$= -\,2{\cdot}7431 \pm 6{\cdot}8846\iota \qquad \text{.........(31)},$$

so that e^{-sz} tends to infinity with depth. These roots are therefore inadmissible. The surface wave given by (25) to (29) is therefore the only one satisfying the conditions.

When $\lambda = \mu$, all the roots of (23) are real; they are

$$c^2/\beta^2 = 4,\ 2\left(1 \pm \frac{1}{\sqrt{3}}\right) \qquad \text{.........(32)}.$$

The smallest of these gives

$$c = 0{\cdot}9194\beta;\ s = 0{\cdot}3933\kappa;\ r = 0{\cdot}8475\kappa \qquad \text{.........(33)},$$

$$u = A\,(e^{-rz} - 0{\cdot}5773e^{-sz}) \sin \kappa\,(x - ct) \qquad \text{.........(34)},$$

$$w = A\,(0{\cdot}8475e^{-rz} - 1{\cdot}4679e^{-sz}) \cos \kappa\,(x - ct) \qquad \text{.........(35)},$$

and the ratio of the axes of the elliptic orbit described by a surface particle is reduced to about 1·5.

The other roots make c greater than both α and β; hence r and s are both purely imaginary. The solution then represents a combination of P and SV waves approaching the surface and being reflected. In neither case is the motion confined to the neighbourhood of the surface.

The surface waves just discussed were discovered by Lord Rayleigh, and are usually named after him. It has been seen in the course of the investigation that no other type of harmonic wave is capable of continuous propagation over the surface of a homogeneous solid; and that in this type the motion is partly vertical and partly horizontal in the direction of propagation, there being no horizontal motion across the direction of propagation. Both compressional and distortional movement is involved.

6·35. Consider now an origin of disturbance some distance below a plane free surface, emitting waves whose length is short compared with the depth of the origin. Then any portion of the wave front, when it reaches the surface, has a radius of curvature large compared with the wave length, and will be reflected like a plane wave. Thus the reflected disturbance will consist of P, SV, and SH pulses, but their relative amplitudes will vary considerably according to the direction of propagation, on account of the peculiarities of reflexion; a spherically symmetrical incident wave does not give spherically symmetrical reflected waves.

If the wave length is comparable with the depth of the origin, the motion is more complicated, because the curvature of the wave front where it meets the surface must now be taken into account. In this case an SH wave is still perfectly reflected as if it came from the image of the origin in the free surface. No combination of the original P or SV wave with reflected P and SV waves, however, will give zero stress over the boundary when the wave-front is curved. But it is found that the conditions can all be satisfied if a system of Rayleigh waves is included. The analysis, which is difficult, was given by Lamb*, and has recently been extended by Nakano†. Several interesting consequences follow at once. In the first place, when a bodily wave is spreading, the original energy becomes spread over a sphere or hemisphere, that is, over an area proportional to r^2, where r is the distance from the origin. Thus the energy density is proportional to $1/r^2$, and the amplitude of the motion to $1/r$. But in a surface wave the depth affected remains the same however far the wave travels, and the energy is spread over an area proportional to r, and therefore the amplitude to $1/r^{\frac{1}{2}}$. Thus when r is great enough the surface wave may produce greater displacements than the bodily wave, even though it may carry only a small fraction of the energy.

When the original disturbance is periodic with wave-length $2\pi/\kappa$, and the depth of the origin is h, it is found that the amplitude of the Rayleigh waves produced contains a factor e^{-rh} if the original disturbance is compressional, and e^{-sh} if it is distortional, where r and s have the meanings given them in 6·34. The deeper the focus, then, the smaller are the Rayleigh waves produced.

6·351. Let us now consider the times of arrival of the waves at a given point. The point on the surface vertically above the origin is called the *epicentre*, while the origin of the disturbance itself is usually known as the *focus*. The distance of the focus below the surface is the *depth of focus*. If now we consider a point Q on the surface distant x from E the epicentre and r from the focus, and the original disturbance is compressional, the P wave will reach Q in time

$$\frac{r}{\alpha} = \frac{(x^2 + h^2)^{\frac{1}{2}}}{\alpha} \qquad \text{.........(1)},$$

the curvature of the earth being neglected.

* *Phil. Trans.* A, 203, 1904, 1–42.

† *Japanese J. of Astron. and Geoph.* 2, 1925, 1–94.

The derived SV wave reaches its maximum in time x/β, and the Rayleigh wave in time x/γ, where γ is the velocity of the Rayleigh waves. These waves begin gradually, and attain their maximum displacements after times equal to the time it would take waves of the respective types to travel from the *epicentre* to the point of observation.

6·352. If the original disturbance is distortional, the S waves reach Q in time $(x^2 + h^2)^{\frac{1}{2}}/\beta$. The P wave in this case has a definite beginning. If reflexion takes place at a point distant ξ from the epicentre, the P wave will reach Q in time

$$\frac{(\xi^2 + h^2)^{\frac{1}{2}}}{\beta} + \frac{x - \xi}{\alpha} \qquad \text{.........(2)},$$

which is a minimum if

$$\frac{\xi}{\beta\,(\xi^2 + h^2)^{\frac{1}{2}}} = \frac{1}{\alpha} \qquad \text{.........(3)}.$$

Thus if $\xi = h \tan i$, so that i is the angle of incidence,

$$\frac{\sin i}{\beta} = \frac{1}{\alpha} \qquad \text{.........(4)},$$

and the shortest time of transmission is therefore determined by the condition that incidence shall be at such an angle as to make the reflected P wave travel along the boundary. This time is found, on substitution in (2), to be

$$\frac{x}{\alpha} + \frac{h \cos i}{\beta}, \text{ or } \frac{x}{\alpha} + \frac{h}{\beta}\sqrt{1 - \frac{\beta^2}{\alpha^2}} \qquad \text{.........(5)}.$$

If then x is great enough for h^2/x to be neglected, S waves will reach Q in time x/β, and thus the P waves will appear to have started later than the S waves by an amount $\frac{h}{\beta}\sqrt{1 - \frac{\beta^2}{\alpha^2}}$. The Rayleigh waves reach their maximum about time x/γ, as in the case of a condensational disturbance.

On the laws of geometrical optics a wave reflected along the boundary would have zero intensity, and the actual P movement arises from diffraction. Nakano in fact finds that its amplitude, like that of the Rayleigh wave, contains a factor e^{-sh}.

It appears, from our outlook on the Rayleigh wave as arising from imperfection of reflexion due to curvature of the wave-fronts, that it is to be regarded as a residual phenomenon. It may, however, be a very large residue. If the solid is such that $\lambda = \mu$, Nakano shows that for a shallow focus 0·417 of the energy sent out from a compressional disturbance, and 0·885 of that from a distortional one, goes into the Rayleigh wave.

Close to the epicentre, incident transformed waves and Rayleigh waves do not exist.

6·4. The threefold character of the disturbance produced by an earthquake, represented by the P, S, and Rayleigh waves predicted theoretically, was first definitely recognized in records of actual earthquakes by R. D. Oldham*, who constructed the first useful tables of the times of transit of the various phases. In several respects, however, the actual movement of the ground during the passage of the waves from an earthquake differs from that indicated by the simple theory. As each phase, for instance, passes a particle, that particle should perform a single complete swing; just as in the case of an explosion wave sent out from a limited region a compressive disturbance is the first to reach any given place, then a rarefaction, after which the whole returns to rest. In an earthquake each phase consists of many oscillations, all nearly simple harmonic. This is not due to an oscillatory character of the original disturbance, for if this consisted of a definite number of oscillations this number would be repeated in the disturbance at every distance from the origin, whereas actually the number of oscillations in each phase ordinarily increases with epicentral distance. Further, the periods are usually different in the various phases. Again, the theoretical motion in a Rayleigh wave is wholly in the plane containing the vertical and the direction of propagation, and the ratio of the vertical and horizontal displacements is about 1·5. In the actual motion in the third phase there is a strong horizontal component at right angles to the direction of propagation. The ratio of the vertical displacement to that in the direction of propagation is usually less than 1·5†; the commonest ratio is 0·9 to 1·0, though values from 1·3 to 1·4 are fairly frequent.

There are three familiar ways of obtaining a train of oscillatory waves as a result of an impulsive disturbance. The first is by the intervention of some mechanism with one or more discrete natural periods, as in the production of a sound wave in air by striking a stretched string. This can hardly arise in the earth's crust: the natural periods form a continuous set and there is nothing to determine which shall be excited. The second is by reflexion of the disturbance at regular intervals. An instance is the high note observed in walking past a set of palings, where the successive boards reflect the sound of one's step at equal intervals so as to give a definite note. The third is dispersion, that is, a dependence of velocity of propagation on frequency. The harmonic waves of varying length sent out by the splash of a stone into water form an example. The second and third causes are probably both relevant in the propagation of seismic waves. In a homogeneous earth P, S, and Rayleigh waves could suffer dispersion only owing to gravity, and the work of Love‡ gives no ground for the belief that this is important in seismology. But also our geological

* *Phil. Trans.* A, **194**, 1900, 135–174.

† C. Mainka, *Phys. Zs.* **16**, 1915, 117–121.

‡ *Problems of Geodynamics*, 1911, Chapter XI.

considerations suggest the need of considering several thin layers of different mechanical properties near the surface; and such layers immediately offer hope of accounting for the chief differences between the seismological facts and the predictions of the theory of waves on a homogeneous elastic solid. Further, they definitely do account for a set of facts not yet mentioned, namely the phenomena of near earthquakes.

6·5. It was found† by A. Mohorovičić, while examining the records of the earthquake of 1909 October 8, in the Kulpa Valley, Croatia, that two distinct compressional and two distortional pulses were present. One pair of these were found to behave, at distances of the order of 1000 km. or more, like the P and S known to previous investigators. The other pair travelled more slowly, but appeared to have started earlier. They were denoted by $\bar{P}$ and $\bar{S}$, but for ease of printing I use instead the notation P_g and S_g. At stations near the epicentre only P_g and S_g were observed. At greater distances P and S arrived before them, so that four distinct pulses could be traced. At still greater distances P_g and S_g could no longer be recognized, but P and S could. P_g and S_g were much larger movements than P and S at stations where all were observable. The interpretation placed on these facts by A. Mohorovičić was that the focus of the earthquake was in an upper layer of the crust, and that P_g and S_g were waves that had travelled in this layer directly from the focus to the observing stations, while P and S had been refracted down into a layer where the velocities of propagation were greater, and afterwards refracted up again. Similar waves were detected in the records of the two Stuttgart earthquakes of 1911 November 16 and 1913 July 20, by S. Mohorovičić‡ and B. Gutenberg§. P_g and S_g, but not P and S, were found in the records of the Oppau explosion of 1921 September 21, by Wrinch and Jeffreys||, Hecker¶, and others. V. Conrad, in discussing the earthquake of 1923 November 28, in the Tauern region of Austria, found the same four waves, and in addition a fifth, which he called P^*, with a velocity of transmission between those of P and P_g††. He suggested that this might be a compressional wave transmitted in an intermediate layer, probably identical with one less definitely observed by Gutenberg in his earlier work. This earthquake and the earlier ones have been rediscussed by the present writer‡‡. In a later paper the existence of the wave P^* was confirmed by finding it

† *Jahrb. d. Meteor. Obs. Agram (Zagreb)*, 1909.

‡ *Beiträge z. Geoph.* **13**, 1914, 217–240; **14**, 1916, 187–198.

§ *Veröff. d. Zentr. Bur. d. Int. Seism. Assn.*, 1915. This is the most comprehensive account of particular near earthquakes yet published.

|| *Observatory*, April 1922, 109; *M.N.R.A.S. Geoph. Suppl.* **1**, 1923, 15–22 (communicated September 1922).

¶ *Veröff. d. Hauptstation f. Erdbebenforschung*, Jena, 1922.

†† *Mitteilungen d. Erdbeben-Kommission*, Wien, 1925, No. 59.

‡‡ *M.N.R.A.S. Geoph. Suppl.* **1**, 1926, 385–402.

in the records of the Jersey and Hereford earthquakes of 1926 July 30 and 1926 August 14†. The corresponding distortional wave S^* was also identified.

So long as the epicentral distance does not exceed about 800 km. all these six pulses seem to travel with uniform velocity; that is, the times of arrival are linear functions of the epicentral distance, apart from small irregular residuals attributable to errors of observation. The velocities found in the writer's discussions were as follows, in kilometres per second.

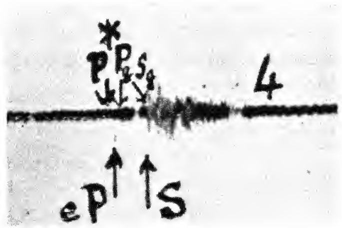
Earthquake	P_g	P^*	P	S_g	S^*	S
Kulpatal	$5{\cdot}6 \pm 0{\cdot}15$	—	$7{\cdot}4 \pm 0{\cdot}3$	—	—	—
Stuttgart, 1911	5·6	—	7·8	—	3·75 (?)	4·2 to 4·3
Oppau	$5{\cdot}4 \pm 0{\cdot}05$	—	—	$3{\cdot}15 \pm 0{\cdot}05$	—	—
Tauern	$5{\cdot}6 \pm 0{\cdot}06$	$6{\cdot}2 \pm 0{\cdot}05$	$7{\cdot}7 \pm 0{\cdot}07$	3·3 (?)	—	—
Jersey	$5{\cdot}4 \pm 0{\cdot}05$	$6{\cdot}3 \pm 0{\cdot}06$	$7{\cdot}8 \pm 0{\cdot}1$	$3{\cdot}3 \pm 0{\cdot}02$	$3{\cdot}7 \pm 0{\cdot}02$	$4{\cdot}35 \pm 0{\cdot}03$
Hereford	$5{\cdot}4 \pm 0{\cdot}04$	—	—	$3{\cdot}3 \pm 0{\cdot}04$	—	$4{\cdot}35 \pm 0{\cdot}04$

The uncertainties given are mean square errors. As a specimen of the accuracy of fit, we may take the Jersey earthquake. Readings and computations were carried out to the nearest second, so that residuals of $\pm$ 1 s. are not significant. For every wave except S_g the majority of the residuals were in this range. For S_g, with times ranging over 204 s., two residuals out of 14 reached $\pm$ 3 s. The distances of the observing stations used ranged up to 942 km. (Hamburg). Wien, at 1300 km., obtained a record, but accurate readings could not be made on account of the smallness of the movement. For the Hereford earthquake and the Oppau explosion the fit is as good, and for the Tauern one even better, but it is tested over a smaller range of distance. The others are less good.

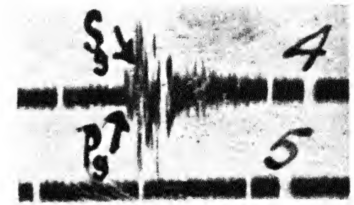
Considering the diversity of the crustal material as revealed by geology, it is remarkable that the propagation of waves in the surface layers should be so uniform. There are signs, indeed, of a difference between western and central Europe in the velocity of P_g, the three disturbances with the most westerly origins having given a velocity of 5·4 km./sec., while the others have given 5·6 km./sec. Since the results are concordant the difference may be genuine; if each of these velocities had been given by only one disturbance it would have remained probable that the velocity was uniform and near 5·5 km./sec., in view of the uncertainties of about 0·05 km./sec. in each velocity, but this seems less likely when the results support one another.

6·51. In reducing the observed times of arrival we find the formula of the form $a + x/v$ that fits the observations of the same pulse at all the stations best. Here x is the distance from the epicentre, and a and v are constants for each pulse: v is the velocity, and a may be called the apparent time of starting. It is found that both v and a are different for the different pulses. This would be expected if the pulses have travelled in different

† *M.N.R.A.S. Geoph. Suppl.* **1**, 1927, 483–494.

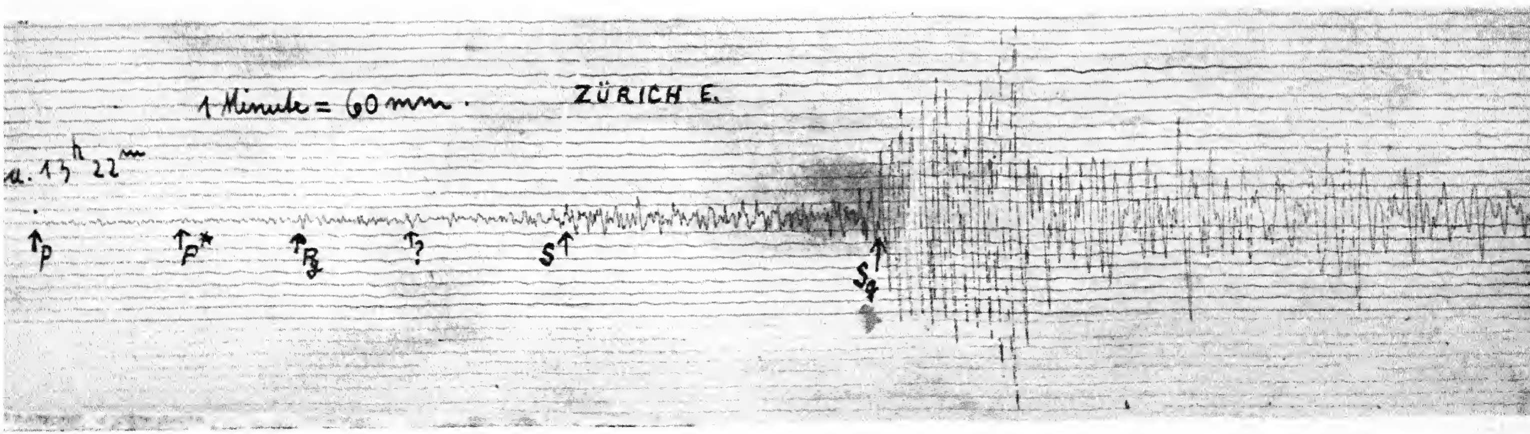


Stonyhurst.



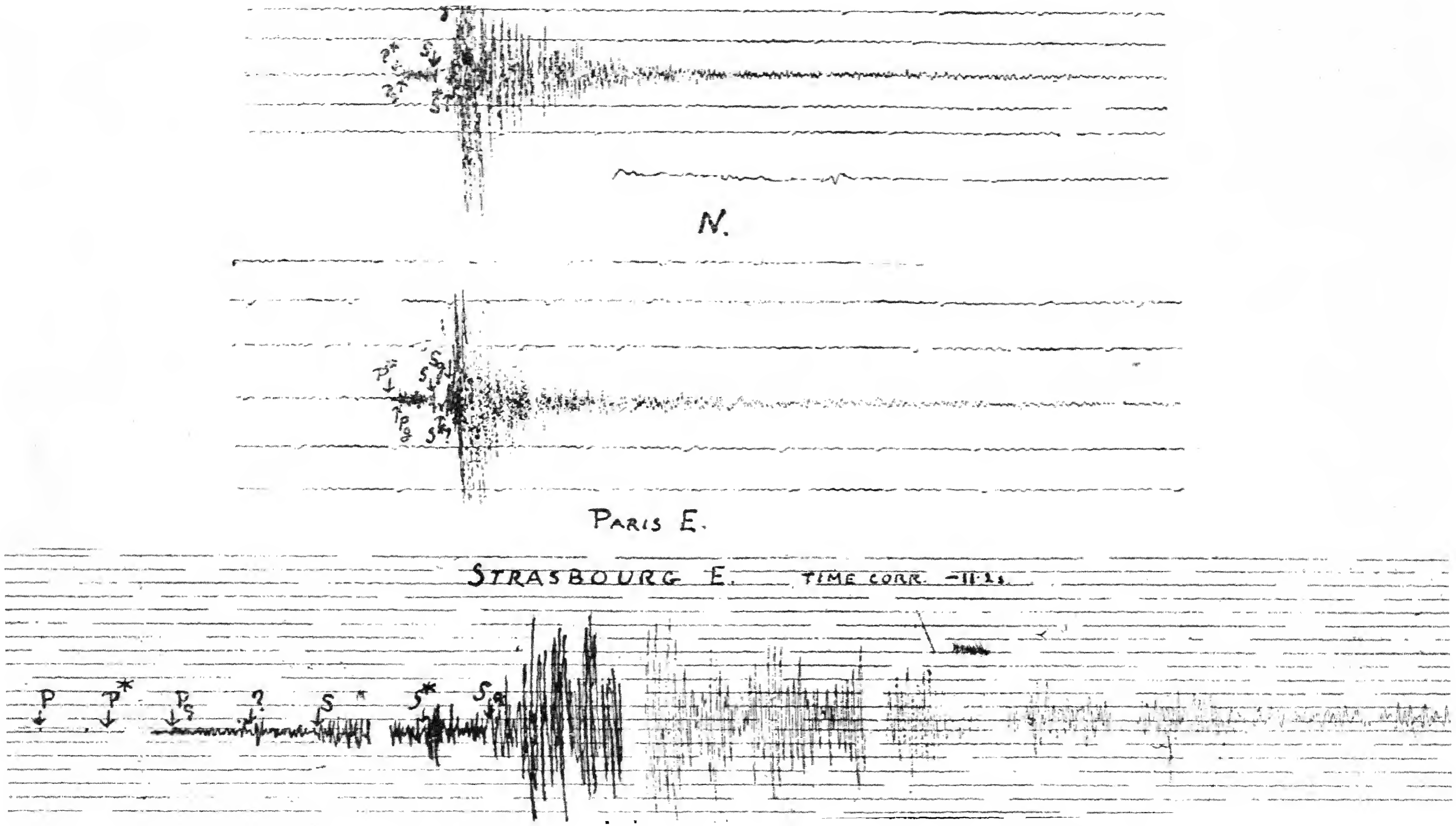
West Bromwich.

Records of Hereford Earthquake. The indications of phases on the Stonyhurst record, shown above the trace, are mine. Those given below the trace were made at the Observatory.

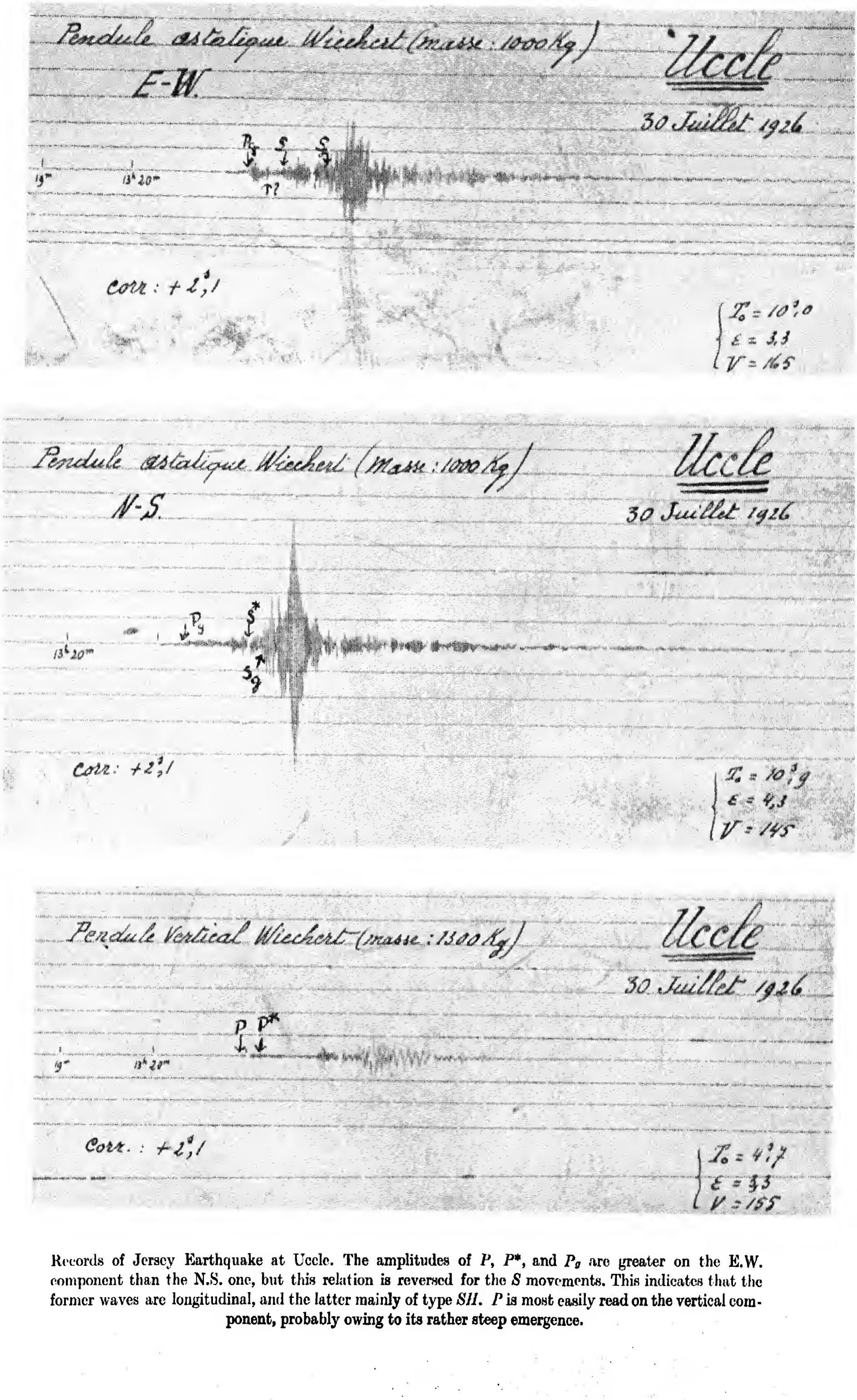


Record of Jersey Earthquake at Zürich. Uccle, Strasbourg, and Zürich all show an unidentified displacement between P_g and S. This may be a compressional wave in the sedimentary layer.

PLATE II



Records of Jersey Earthquake at Paris and Strasbourg. On the Paris record, P^*, which is a very small and rapid oscillation, too small to reproduce clearly, has been emphasized by thickening the trace by hand on a reproduction. On the Strasbourg one, also, it has been found desirable to ink in the trace for two minutes.



Records of Jersey Earthquake at Uccle. The amplitudes of P, P^*, and P_g are greater on the E.W. component than the N.S. one, but this relation is reversed for the S movements. This indicates that the former waves are longitudinal, and the latter mainly of type SH. P is most easily read on the vertical component, probably owing to its rather steep emergence.

layers, for delays would arise owing to the time spent in travelling up or down from one layer to another. The time of transit from the focus to the observing station is in any case

$$T = \int \frac{ds}{c} \qquad \text{.........(1)},$$

taken along the actual path, where c has of course the value at each point appropriate to the type of the wave there. But if we take horizontal and vertical coordinates x and z, and introduce i, the angle of incidence on the horizontal surface through a point of the path, we have

$$dx = \sin i\, ds; \; dz = \cos i\, ds; \; T = \int \frac{\operatorname{cosec} i}{c}\, dx \qquad \text{.........(2)}.$$

On the other hand if c' is the velocity at the lowest point of the path, the time calculated on the hypothesis that the wave travels the whole distance x with velocity c' is

$$T' = \frac{x}{c'} = \int \frac{1}{c'}\, dx \qquad \text{.........(3)}.$$

But by the laws of refraction, since i is 90° at the lowest point,

$$\sin i = c/c' \qquad \text{.........(4)},$$

whence, by subtracting (3) from (2) and substituting for c, we have

$$T - T' = \int \frac{\cot^2 i}{c'}\, dx = \int \frac{\cot i\, dz}{c'} = \int \frac{1}{c}\left(1 - \frac{c^2}{c'^2}\right)^{\frac{1}{2}} dz \qquad \text{.........(5)}.$$

The horizontal part of the path contributes nothing to this integral. As the ray goes down and up again, we have

$$T - T' = \int_h^H \frac{1}{c}\left(1 - \frac{c^2}{c'^2}\right)^{\frac{1}{2}} dz + \int_0^H \frac{1}{c}\left(1 - \frac{c^2}{c'^2}\right)^{\frac{1}{2}} dz \qquad \text{.........(6)},$$

where h is the depth of the focus and H that of the lowest point of the ray. If h is equal to H, the first of these integrals vanishes; if h is 0, they are equal, and $T - T'$ is double the second. When the velocities are known $T - T'$ gives a means of determining the thicknesses of the layers.

We notice first that the influence of the curvature of the earth is unimportant in the study of near earthquakes. We are not concerned with epicentral distances over 800 km. or so, and an arc of 800 km. subtends at the centre of the earth an angle whose circular measure is 0·13. The arc differs from the chord by just over 1 km., so that the times of transmission of S_g, the slowest pulse, along the arc and the chord, would differ by only a third of a second. As the effect is proportional to the cube of the distance it is clear that it is within the limits of the residuals.

In the linear formula for the time of arrival of a wave, v is the same as c', the velocity on the horizontal part of the path, and x/v is therefore T'. $T - T'$ is the difference between the apparent and true times of starting, and may be called the 'delay in starting.' The true time of starting is

the time of the original fracture, and is equal to the constant a in the formula appropriate to S_g. The delays in starting for the Jersey earthquake were found to be as follows:

$$S_g,\ 0;\ P_g,\ 3\,\mathrm{s.};\ P^*,\ 5\,\mathrm{s.};\ P,\ 9\,\mathrm{s.};\ S^*,\ 4\,\mathrm{s.};\ S,\ 8\,\mathrm{s.}$$

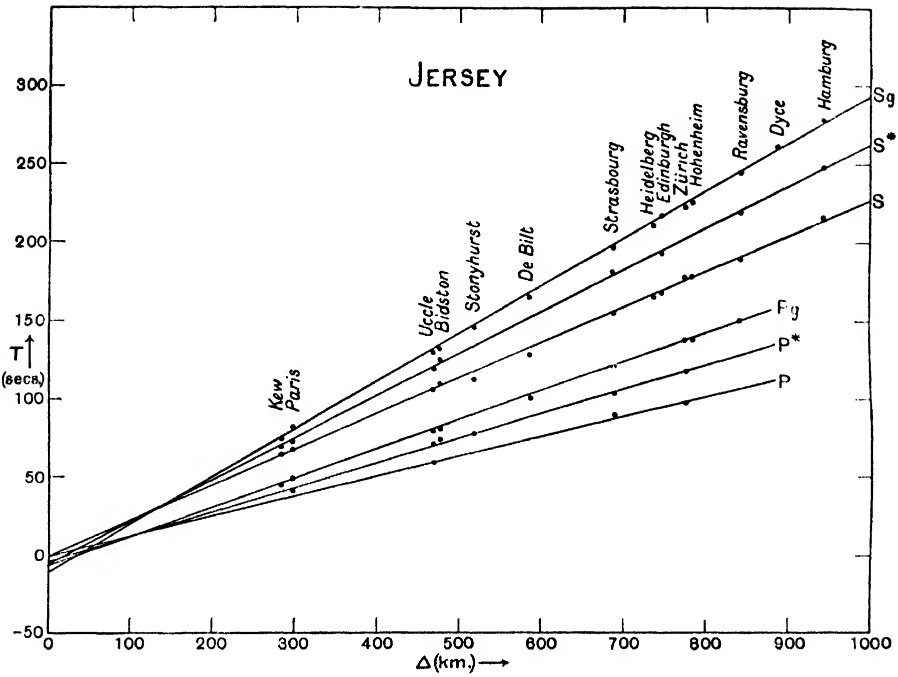
For the Hereford one the delays of P_g and S in comparison with S_g were respectively 2 s. and 9 s.; the others were not well determined. In the Stuttgart one of 1911 the delay of P in comparison with P_g was 6 s. In the Kulpatal one, when the straightforward least square solution was adopted, this delay was found to be only 3·6 s. But the velocity found for P was lower and less definitely determined than the other earthquakes indicate; if we adopt instead the value 7·7 km./sec. the delay is found to be 7 s. In the Tauern earthquake the delays of P^* and P with reference to P_g were 2·0 s. and 6·9 s.

In the Jersey and Hereford earthquakes P_g seems to have started 2 or 3 seconds after S_g. This is a curious result, but seems to be genuine. To explain it as error of observation we should have to assume that P_g was read systematically late, or S_g systematically early, by two or three seconds on every record, which appears impossible. The foci were in any case in the upper igneous layer, and there might be delays when the pulses passed through the overlying sedimentary layer. But a compressional wave would certainly travel faster in this layer than a distortional one, so that S_g would appear to be delayed with reference to P_g, whereas the contrary is observed. The most probable explanation seems to be as follows. The original disturbance would certainly be a fracture, and fractures occur as a means of relieving distortional stress. It is therefore probable that the fundamental wave sent out from an earthquake is usually mainly or wholly distortional. But if so, this wave is S_g, and P_g must be derived from it by reflexion at the outer surface or the base of the sedimentary layer. Then P_g will show a delay in starting, according to 6·352 (5), amounting to $\frac{h}{\beta}\sqrt{1-\frac{\beta^2}{\alpha^2}}$, where α and β are now the velocities of P_g and S_g. If h is in kilometres this delay is $h/3{\cdot}9$ seconds. The delays found would therefore imply that the depths of foci of the Jersey and Hereford earthquakes were about 12 km. and 8 km. respectively.

A similar phenomenon has since been noticed in the Montana earthquake of 1925 June 28†.

If the origin is in the upper layer the waves P^* and P must be derived from P_g; and if the latter is itself derived wholly from S_g, the delay of P_g arising from the depth of focus will affect P^* and P equally. The delay of P^* with reference to P represents in fact two passages (one up and one down) through the upper layer, and that of P with reference to P^* two through the intermediate layer. We can with sufficient accuracy treat

† *M.N.R.A.S. Geoph. Suppl.* **1**, 1928, 504.



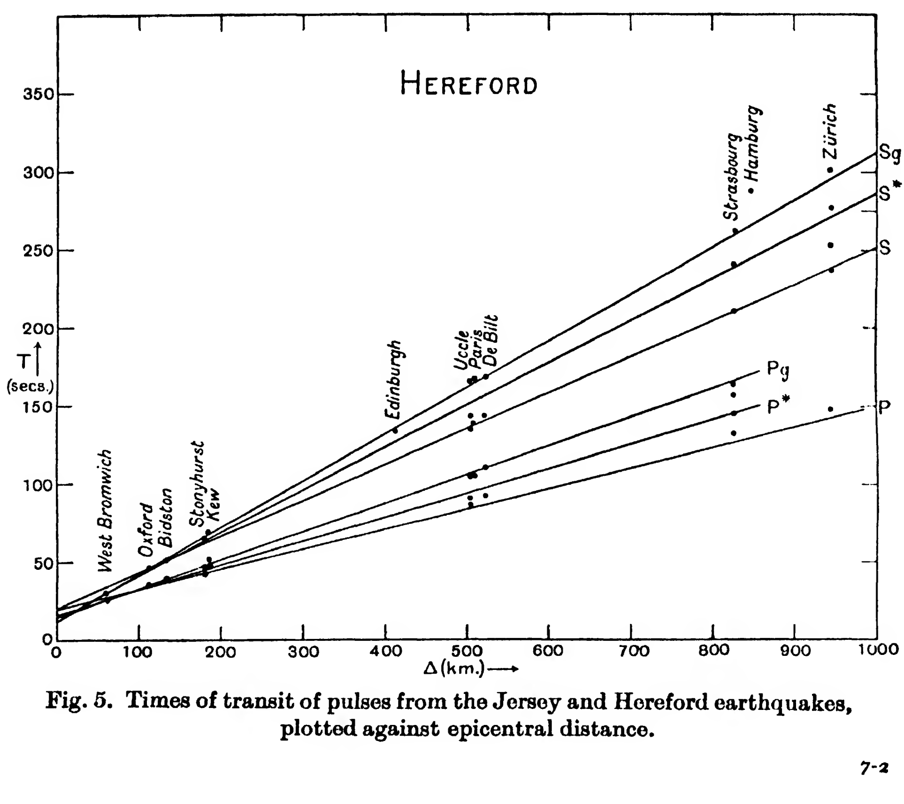


Fig. 5. Times of transit of pulses from the Jersey and Hereford earthquakes, plotted against epicentral distance.

both layers as homogeneous. Applying (6), we have, for the delay of P^* relative to P_g, $\frac{1}{8}$ s. for each kilometre of depth of the upper layer. Hence a delay of 2·0 s. implies that this layer is 10 km. thick. The delay of P relative to P^* is $\frac{1}{5 \cdot 4}$ s. for each kilometre of depth of the intermediate layer. Our estimates of 4 s. from the Jersey earthquake and 4·9 s. from the Tauern one therefore imply thicknesses of 22 km. and 26 km. respectively. The results from the Stuttgart and Kulpatal earthquakes are consistent with these.

The waves S^* and S are derived from S_g directly. Their delays in starting can therefore be calculated from the velocities, the thicknesses of the layers, and the depth of focus. In this case the vertical distance

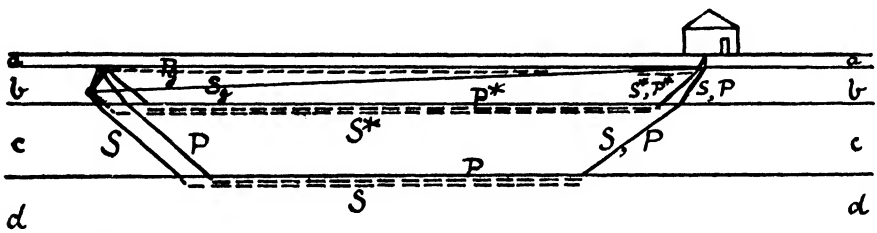


Fig. 6. Diagram of the probable paths of the six pulses observed in near earthquakes. Broken lines indicate waves propagated along or near to boundaries. The horizontal scale is, of course, much smaller than the vertical; the angles are approximately correct. *aa*, Sedimentary layer; *bb*, granitic layer; *cc*, intermediate layer (basalt or diorite); *dd*, lower layer (dunite, peridotite, or eclogite). The compressional waves may actually be generated at the upper surface, and not at the bottom of the sedimentary layer, as shown in the diagram. The transitions also may be gradual and not sudden.

traversed is twice the depth of the lowest point reached, less by the depth of focus. When these delays were calculated for the Jersey and Hereford earthquakes the results were as follows, in seconds:

	Jersey		Hereford
	S^*	S	S
Calculated	1·3	7–8	8
Observed	4	8	9

The agreement is as good as could be expected, since S^* is the most difficult of the six waves to observe accurately†.

6·52. *Constitution of the Crustal Layers.* The velocities of elastic waves in the upper layers of the crust give some information about the elastic properties of the materials. If this information is compared with experimental determinations of the properties of the rocks known from geology to be probable, we can proceed a long way towards identifying the main constituents of the crust. The velocities, however, are difficult to measure

† Consistent results for Californian earthquakes have been obtained by P. Byerly, *Bull. Seism. Soc. Amer.* **17**, 137–146, 203–206, 213–217.

accurately in the laboratory, and the most accessible elastic property is the compressibility; that is, the fractional change of volume per unit increase of hydrostatic pressure. In terms of the constants λ and μ, the bulk modulus k is $\lambda + \frac{2}{3}\mu$, and the compressibility is $1/k$. From our formulae for the velocities, 6·3 (5) and (8), we have

$$k = \rho\,(\alpha^2 - \tfrac{4}{3}\beta^2) \qquad \text{.........(1).}$$

Now our seismological results give α and β, and the corresponding values of $\alpha^2 - \frac{4}{3}\beta^2$ are, in c.g.s. units, as follows:

	$\alpha\,(\times 10^{-5})$	$\beta\,(\times 10^{-5})$	$\alpha^2 - \frac{4}{3}\beta^2\,(\times 10^{-10})$
Upper layer	5·4	3·3	14·7
	5·6	3·3	16·9
Intermediate layer	6·3	3·7	21·4
Lower layer	7·8	4·35	35·6

The compressibilities of rocks have been investigated experimentally by several workers, especially L. H. Adams, E. D. Williamson, and R. E. Gibson of the Geophysical Laboratory at Washington. For mixed rocks the most reliable way of finding them at high pressures is to determine the compressibilities of the component minerals separately and form an average weighted according to the known composition by volume. In the work the changes of volume of the specimens under pressure were compared with those of a standard cylinder of soft steel, and the known compressibility of this was used to infer those of the rocks. In the work of Adams and Williamson* a representative series of rocks was tested. The compressibility of the standard piece of steel, however, was given too high a value, and all their compressibilities need to be reduced by $0{\cdot}02 \times 10^{-12}$ per c.g.s. unit of pressure in consequence. This has been done in the table on p. 102. In a later paper† Adams and Gibson determined the values for dunite and tachylyte. In the table, where two rows of results are given for a rock, the upper row corresponds to a pressure of 2×10^9 c.g.s. and the lower to one of 10^{10} c.g.s., corresponding to depths of 7 km. and 33 km. respectively in the crust, together with average densities. Where only one compressibility is given, it is a mean over the range between 2×10^9 and 10^{10} c.g.s. The pyroxenite considered was half augite and half hypersthene, the pallasite half olivine and half metallic iron, and the siderite all metallic iron.

The bulk-moduli given are considerably higher than those found by previous workers, who used lower pressures. Thus F. D. Adams and E. G. Coker‡ got for several granites values of $1/k$ of $3{\cdot}0$ to $3{\cdot}6 \times 10^{-12}$, for New Glasgow gabbro (two specimens) $1{\cdot}18$ and $2{\cdot}13 \times 10^{-12}$, and for a syenite $2{\cdot}33 \times 10^{-12}$. It appears that when the pressure is increased from one atmosphere the compressibility is great at first, but falls off before

* *J. Frank. Inst.* **195**, 1923, 475–529.

† *Proc. Nat. Acad. Sci.* **12**, 1926, 275–283.

‡ *Amer. J. Sci.* (4) **22**, 1906, 95–123.

the pressure has reached 2×10^9 c.g.s. and varies much more gradually at higher pressures. In seismology we are concerned mainly with the higher pressures, and it appears that those employed at the Geophysical Laboratory cover the range of conditions concerned in the study of near earthquakes. On the other hand the experimental work has, for obvious reasons, been done at ordinary temperatures. We shall expect the elastic constants within the earth therefore to be somewhat lower than the experimental values on account of the higher temperatures involved.

Rock	Density	$1/(k \times 10^{-12})$	$k \times 10^{-12}$	$k/\rho \times 10^{-10}$
Granite	2·61	2·10	0·476	18·3
	2·66	1·86	0·538	19·5
Granodiorite	2·69	1·81	0·552	20·5
	2·73	1·64	0·610	22·3
Syenite	2·61	1·85	0·540	20·7
	2·66	1·66	0·602	22·7
Diorite	2·74	1·60	0·625	22·8
	2·78	1·47	0·680	24·3
Gabbro	3·05	1·18	0·847	27·8
	3·08	1·15	0·870	28·2
Pyroxenite	3·40	1·01	0·990	29·0
	3·44			
Peridotite	3·40	0·95	1·05	30·7
	3·44			
Dunite	3·29	0·84	1·19	35·2
	3·32	0·79	1·27	38·3
Pallasite	5·65	0·75	1·33	23·5
	5·69			
Siderite	7·9	0·58	1·72	25·0
Obsidian	2·333	2·86 *	0·350	15·1
Tachylyte	2·851	1·45*	0·690	24·2

Now comparing $\alpha^2 - \frac{4}{3}\beta^2$ in the table on p. 101 with k/ρ in this table, we see that the value for the upper layer of seismology is below that for granite, and near that for obsidian at high pressures. The granitic layer definitely exists, and the bulk modulus in it must be below the laboratory value at the same pressure, on account of the higher temperature. This consideration seems to rule out obsidian and to indicate the extent of the effect of temperature within the granitic layer. Again, the only rock that is known to give the requisite value for k/ρ in the lower layer is dunite. If then we identify the lowest layer with dunite, the basaltic layer must be the intermediate one. But remembering that basalt, gabbro and tachylyte are chemically the same, and that mere fineness of texture would not affect the elastic properties, we notice that tachylyte fits the facts much better than gabbro; in fact we may say that

* Bridgman (*Amer. J. Sci.* **10**, 1925, 359–367) gives about 2·5 (variable) for obsidian, 1·87 for tachylyte from Torvaig, Skye, and 1·33 for tachylyte from Kilauea.

the seismological evidence is inconsistent with the presence of any widespread layer of crystalline basaltic matter in the crust*.

There are, however, doubts about this interpretation. Some basalt near the surface has been metamorphosed by stress into eclogite, and consequently it has come to be widely believed that the basalt at any considerable depth in the crust is in the form of eclogite. The elastic properties of this rock are not known, but if it has a value of k/ρ agreeing with that of the lower layer, as seems not unlikely, the intermediate layer could be interpreted as diorite. This alternative was advanced by Holmes†. On the other hand Bridgman finds some abnormalities in the compressibility of tachylyte, which lead Daly‡ to consider seriously the possibility that the velocities in the lower layer may fit tachylyte after all. But this seems to me to stretch the data too far. These ideas raise questions to be settled by further experimental work. At the same time the possibility that the lower layer is actually dunite in a vitreous state is not excluded. If the upper layer is granite and the intermediate one tachylyte, rocks of dioritic composition must be considered as derived from a transition zone between these two layers.

6·53. A different method of discussing observations of near earthquakes has been given by Prof. S. Mohorovičić in two papers already mentioned. It is obvious that for a focus at a depth h the time of transmission of P_g to a station with epicentral distance x would be $(x^2 + h^2)^{\frac{1}{2}}/a$, where a is the velocity of P_g. When x/h is great this approximates to x/a and gives a linear relation. But at epicentral distances comparable with h the time of transit will exceed x/h; if the time is plotted vertically and the distance horizontally, the resulting curve will be concave upwards for short epicentral distances. On the other hand, when the epicentral distance is great the linear relation will not be exact for two other reasons. Even in a homogeneous crust the time of transmission will increase rather less rapidly than the epicentral distance measured as an arc, since the wave will travel along the chord. This will be accentuated in a layer where the velocity of propagation increases with depth, because the wave economizes time further by travelling at a greater depth than the chord. Consequently the graph of the time of transmission against x should be concave upwards for small values of x, and concave downwards for large values. Prof. Mohorovičić shows that in a non-uniform crust the point of inflexion corresponds to the ray that leaves the focus horizontally. If the curve was accurately known, then, it would be possible to use its departures from linearity to find both the depth of focus and the variation of velocity with depth. This has been attempted by Prof. Mohorovičić, who claims to have obtained definite results; they are inconsistent with those obtained

* Jeffreys, *Nature*, **118**, 1926, 443.

† *Nature*, **118**, 1926, 586.

‡ *Amer. J. Sci.* **15**, 1928, 108–135.

in my discussions and with those of several other lines of investigation of the structure of the crust. I consider them to arise from inadequacy of the observational material.

We have seen that it is the exception and not the rule for an observation of the time of P_g to depart from the simple formula $a + x/v$ by more than a second. Estimates of the velocity depend on the accuracy of this formula; estimates of the depth of focus and the variation of velocity with depth depend on its being inaccurate. Departures from linearity are in any case small. The question at issue is whether the data are capable of determining them. This is a matter of the theory of combination of observations. Prof. Mohorovičić's curve for the times of transmission of P_g from a superficial origin up to a distance of 700 km. nowhere deviates from a straight line by more than 0·23 s. Of the three disturbances he specially considers in a reply to my discussion*, namely the two Stuttgart ones and the Oppau explosion, the curves obtained by simply smoothing the data would deviate in the middle from a straight line by, respectively, + 6, − 5, and + 6 times as much as his curve does†; but a statistical discussion shows that the smoothed curves themselves are uncertain by amounts comparable with their departures from the straight line. All we can say is that the linear laws for waves from a superficial origin are correct within about a second, and that we have no means of determining whether there is any significant systematic departure from them.

The question of depth of focus arises from the fact that in three cases, the Kulpatal earthquake and the two Stuttgart ones, the time of P_g at the station nearest to the epicentre differed from the formula by + 3 to + 5 seconds. In the absence of other evidence it would be natural to attribute these residuals to depth of focus. But other evidence, even in the records of the same earthquakes, turns out to be inconsistent with the estimates of depth of focus that they lead to. Let us consider in particular the two Stuttgart disturbances. The times of P_g are given in Gutenberg's memoir, mostly to the nearest second. In the columns $O - C$ of the table on p. 105 those for the nearer stations have been compared with times calculated on the hypothesis of a superficial focus, the times of the original shocks being taken respectively as 25 m. 50 s. and 6 m. 20 s. respectively, hours being omitted.

The times of transmission in seconds have been taken as $\Delta/5{\cdot}6$, where Δ is the epicentral distance in kilometres. The ordinary residuals in such discussions being about $\pm$ 1 s., there is no occasion for comment in any of the observations except those at Hohenheim. But if we accept the two observations at Hohenheim we must say that the distances of the foci from this station exceeded the epicentral distance by 4·9 and 3·9 seconds' journey respectively, that is, by 27·5 and 22 km. The focal depths needed

* *Gerlands Beitr. z. Geoph.* **17**, 1927, 180–231.

† *Ibid.* 417–427.

to account for these would be 57 km. and 50 km. respectively. But these would affect the times of arrival at other stations. The last two columns of the table give, for comparison, the delays introduced by depth of focus at all the other stations as far as München, from the formula

$$R = \{(\Delta^2 + h^2)^{\frac{1}{2}} - \Delta\}/5{\cdot}6,$$

the column R_1 corresponding to $h = 50$ km. and R_2 to $h = 35$ km. If either of these focal depths were correct for either earthquake the corresponding column R_1 or R_2 should agree closely with the column $O-C$ for that earthquake. Actually we see that, whereas without allowance for focal depth only one residual for each earthquake exceeds 2 s., the column R_1 now gives residuals of 2 s. and over at Biberach, Zürich (?), Nördlingen,

Station	Δ (km.)	$O-C$ (I)	$O-C$ (II)	R_1	R_2
Hohenheim	45	+ 4·9	+ 3·9	3·9	2·1
Biberach	56	− 1·0	+ 1·0	3·4	1·8
Durlach	89	—	− 0·9	2·3	1·3
Karlsruhe	94	+ 1·2	+ 1·2	2·2	1·1
Freiburg	100	—	+ 0·1	2·1	1·1
Strassburg	104	+ 1·0	+ 0·4	2·0	1·1
Zürich	113	− 0·2 to + 0·8	− 0·2	2·0	0·9
Nördlingen	119	− 1·2	− 0·2	1·8	0·9
Heidelberg	123	− 1·0	+ 2·0	1·8	0·9
Jugenheim	164	− 1·3	− 1·3	1·4	0·7
München	185	− 0·3	+ 1·0	1·3	0·6

Heidelberg and Jugenheim for the first earthquake, and Biberach, Durlach, Freiburg, Zürich, Nördlingen, and Jugenheim for the second. The column R_2 gives residuals of 2 s. and over at Hohenheim, Biberach, Nördlingen, and Jugenheim for the first earthquake, and Durlach and Jugenheim for the second. Further, while the residuals on the hypothesis of zero depth of focus are unsystematic, the allowance for depth of focus makes them nearly all negative. The only conspicuous exception is Hohenheim, which retains residuals + 1·0 s. and 0·0 s. on the hypothesis that the depth was 50 km., and + 2·8 s. and + 1·8 s. for depth 35 km. The former pair would be satisfactory if the previous agreement at several other stations was not spoiled at the same time; the latter show that depth 35 km. does not even fit satisfactorily the observations at Hohenheim, which were the only reason for assuming considerable focal depth in the first instance.

Further evidence is provided by the arrival of S_g at Hohenheim and Biberach. The interval between P_g and S_g was about 6 s. at both stations for both earthquakes. The differences between the times taken by these waves to travel 45 km. and 56 km. respectively would be 6 s. and 7·5 s. There is no indication here that the focal depth was more than a few kilometres.

Again, 6·51 (6), applied to the delay of P^* with reference to P_g in the Tauern and Jersey earthquakes, was found to lead to an estimate of about

10 km. for the thickness of the granitic layer. If we follow Mohorovičić and Gutenberg and suppose P_g to come direct from the focus instead of being generated by reflexion of S_g, and if further we take the depth of focus equal to the thickness of the layer, the estimate of the thickness can be doubled, but no more. The delay of P^* is inconsistent with any thickness of the granitic layer over 20 km.; and the focus cannot be 35 km. deep in a layer that at the most is only 20 km. thick.

Finally, the view that P_g is derived from S_g and is not a primitive wave is demanded by the observations of the Jersey earthquake at 14 stations, and of the Hereford one at 10, and the evidence of these earthquakes requires the thickness of the upper layer to be about 10 km.

S. Mohorovičić in his early discussion estimated the depth of focus at about 35 km. without considering whether the Hohenheim residuals could have been produced by some cause other than depth of focus, as the above independent lines of evidence all indicate. Gutenberg, in his memoir, says (pp. 64, 65): "Thus it is seen that on the ground of the time-observations of the first earthquake no conclusion concerning the depth of focus is possible"; "The departures of the calculated curves (involving focal depths from 4 km. to 45 km.) from one another lie within the errors of observation, and a determination of the most probable depth of focus is impossible." Finally he says (p. 66): "It is impossible to decide on a depth of focus from time-observations, so long as there are no observations in the immediate neighbourhood of the focus trustworthy to 0·1 or 0·2 s." Gutenberg, however, has abandoned these perfectly sound views without answering his own arguments, and has recently accepted† a depth of about 35 km. with the remark: "That H. Jeffreys has found an appreciably smaller value for these earthquakes arises from the fact that he uses only the first earthquake and attaches special weight to the arrival at Biberach, which is uncertain by at least 3 s." It has been shown above that the observation at Biberach was only one out of eight that had negative residuals when the depth of focus was chosen to fit Hohenheim; I myself actually called attention to its uncertainty when it had not been mentioned by Mohorovičić. The second earthquake is discussed above. Whether the residuals of + 1·8 s., − 0·8 s., and − 2·2 s. at the nearest stations, given by a depth of focus of 35 km., are appreciably more satisfactory than my + 3·9 s., + 1·0 s., and − 0·9 s. is perhaps a matter for opinion. Gutenberg does not discuss any of the other evidence. It may be mentioned that the uncertainty that affects the time at both Hohenheim and Biberach does not extend to the intervals between P_g and S_g, and that these speak definitely for a small focal depth.

The above arguments for a considerable focal depth have been discussed at length because they form an essential part of the reasons advanced by Mohorovičić and now by Gutenberg for adopting a thickness

† *Gerlands Beitr.* **18**, 1927, 379–382.

of about 57 km. for the upper layer or the two upper layers. The need for a much smaller thickness is fundamental in so many of the problems treated in this book that a full analysis of any arguments opposed to it has been considered necessary.

6·54. The amplitude of an elastic wave spreading out symmetrically from a focus should be inversely proportional to the distance. The observed amplitudes of P_g and S_g are more nearly inversely proportional to the epicentral distance Δ than to any simple power of it. This is very good so far as it goes. For waves of type *SH* it is almost the whole story; for when these waves are reflected at a free surface the reflected waves make the same contribution to the motion of the surface as the incident ones, and the amplitude remains proportional to $1/\Delta$. But the proportionality also seems to hold for P_g and S_g, and for both of these the motion of the ground, for a given amplitude and small angle of emergence†, is proportional to the angle of emergence and therefore to $1/\Delta$. Thus the motion of the ground should be proportional to $1/\Delta^2$ instead of to $1/\Delta$. I think the amplitude in the bodily P_g and S_g waves must really be nearly proportional to $1/\Delta$, as the work of Lamb and Nakano implies, but that these waves are brought up to the surface at moderate angles of emergence by refraction at the possibly irregular base of the sedimentary layer. If so, the angle of emergence at the outer surface will on the average be nearly independent of Δ, and the amplitudes will be proportional to $1/\Delta$, as the observations seem to require.

For the indirect waves the question is still more complicated. Where all the six waves are shown on the record, the order in time is ordinarily also the order of increasing size, namely P, P^*, P_g, S, S^*, S_g, though exceptions occur. But each phase initiates a series of oscillations, and the timing of the next phase involves distinguishing the new pulse from the wave motion still surviving from the previous phases. So long as each successive impulse is larger than the previous ones, and has a sharp beginning, this is not very difficult. If these conditions do not hold, however, reading may be difficult; these troubles are particularly common with S^*, which often seems to grow quite continuously out of S. Ordinarily P is about $\frac{1}{10}$ as large as P_g, and like it its amplitude varies irregularly, but on the whole like $1/\Delta$. This is difficult to explain. The variation of amplitude with Δ can be calculated on the assumption that the laws of geometrical optics hold; it is found that while at epicentral distances greater than about 3000 km. the calculated and observed amplitudes are reasonably consistent, the correspondence breaks down completely at short distances‡. The calculated amplitude is *directly* proportional to Δ at these distances: this arises mainly from the fact that the wave is refracted into the lower

† The angle of emergence is the complement of the angle of incidence.

‡ *M.N.R.A.S. Geoph. Suppl.* **1**, 398–401.

layer at nearly grazing incidence and therefore transmission is small. Even if the reduction of transmission is not allowed for agreement is still not obtained. To account for the actual comparatively large size of P it is necessary that the laws of geometrical optics shall cease to hold at short epicentral distances; the indirect waves are then to be considered as diffracted around the boundaries. This hypothesis, when examined theoretically, gives a considerable improvement†, but the amplitude of P now varies as $1/\Delta^2$ instead of $1/\Delta$. The explanation of the variation awaits further investigation.

Another peculiarity probably related to the behaviour of the amplitudes is that P and P^* do not appear as near the epicentre as we should expect. On account of the larger size of P_g and S_g these swamp the indirect waves wherever the distance Δ is short enough for P_g and S_g to arrive first. But when we adopt the empirical formulae that fit the times of P_g, P^*, and P at the distances where they are observed, we find that P^* should arrive before P_g at distances over 80 km., P before P_g for $\Delta > 100$ km., and P before P^* for $\Delta > 130$ km. Actually the less direct waves are not observed at distances anywhere near these lower limits: they first appear at such distances that they already precede the more direct ones by several seconds. Thus in the Jersey earthquake P was not identifiable before P^* at Paris, distant 297 km., though in the Tauern one it was detectable at Innsbruck, distant 169 km., and in the first Stuttgart one at Jugenheim, distant 164 km., but not at Heidelberg, distant 123 km. It looks as if the diffraction that is needed to account for the large amplitudes of P where it appears does not work at these comparatively short distances.

The waves P_g and S_g are not observable at great epicentral distances. The greatest distance where the former has been detected was at Ravensburg, which got it from the Jersey earthquake at $\Delta = 841$ km. It seems to be gradually damped out as it advances, while S_g is apparently converted into surface waves.

6·55. A complete determination of the energy of an earthquake would be difficult, but a rough estimate can be made fairly easily. We notice first that a moderate fraction of the original energy would be in the form of SH waves; it is usually found that in the distortional phases the horizontal motion is mainly across the direction of propagation. The observed velocity of the ground gives the kinetic energy per unit volume; on an average the potential energy is the same. The length of the part of the wave train carrying most of the energy, namely the large waves following close on S_g, is the product of the velocity of the waves composing it and the time they take to pass over a given place, and the energy per unit volume may be taken to be the same throughout the granitic layer. Finally we multiply by $2\pi\Delta$ to get the volume of the region occupied by the wave train

† *Proc. Camb. Phil. Soc.* **23**, 1926, 472–481.

when its middle was passing the station. Multiplying this by the total energy per unit volume we have the energy of the main part of the waves. The whole energy of the earthquake is unlikely to be more than about twice this. In this way it is found that the energy of the Hereford earthquake was about 5×10^{16} ergs; the Oppau explosion was comparable with this; the Jersey earthquake about 10^{19} ergs; the Montana one about 10^{21} ergs. The Pamir earthquake of 1911 February 18 was found, mainly from a consideration of the Rayleigh waves sent out, to have had an energy of about 10^{21} ergs*. These estimates may be compared with the greatest distances where the respective disturbances were fully observed with ordinary instruments, with magnifications up to 250. The Oppau explosion was thus observed at 365 km., the Hereford earthquake at 522 km., the Jersey one at 942 km., and the Montana and Pamir ones at distances over 11,000 km.

With regard to the Oppau explosion, it may be mentioned that it was produced by the explosion of 4500 tons of the double salt $2NH_4NO_3 . (NH_4)_2SO_4$. Assuming that the heat of decomposition of ammonium nitrate is 7500 calories per gram molecule†, the energy liberated must have been about $1{\cdot}5 \times 10^{12}$ calories or 6×10^{19} ergs. Only a small fraction of this went into the seismic waves. The impulses upward on the air and downward on the ground must, by Newton's third law, have been equal. Hence the energies imparted to the earth and the air must have been in the ratio of the initial vertical velocities, which must have been hundreds or thousands of times greater for the air than for the ground. Thus only a small fraction of the energy of the explosion could have entered the earth; the rest went into the sound wave in the air.

6·6. The differences between the surface waves found on the theory of disturbance within a homogeneous solid, and the surface waves observed in actual earthquakes, have already been mentioned (6·4). They largely disappear when the effects of heterogeneity of the crust are considered. Love showed that when a layer of finite thickness rests on a deep solid layer, surface waves could exist such that the displacement is everywhere horizontal and at right angles to the direction of propagation. It was necessary that the velocity of distortional waves in the lower layer should exceed that in the upper; this condition is satisfied by the velocities found in the investigation of near earthquakes. Subject to this condition these transverse waves can proceed any distance without loss of energy, and therefore provide a natural explanation of the displacements at right angles to the plane of propagation observed in the surface waves of earthquakes.

Love also found that in a heterogeneous crust both these waves and Rayleigh waves are subject to dispersion. For Rayleigh waves this is

* Jeffreys, *M.N.R.A.S. Geoph. Suppl.* 1, 1923, 22–31.

† I am indebted to Dr E. K. Rideal for this information.

obvious. For the amplitude of the motion of these waves dies down with depth in a distance proportional to the wave length. Consequently in a crust of two homogeneous layers the movement in very short waves is practically confined to the upper layer, and the velocity is therefore that characteristic of simple Rayleigh waves in the upper layer; while very long waves extend so deeply into the lower layer that the upper layer hardly affects them, and therefore have the velocity that Rayleigh waves on the lower layer would have if the upper layer were absent. Since dispersion ordinarily converts an impulse into a train of approximately harmonic waves the oscillatory character of the third phase is qualitatively explained.

6·61. *Love Waves.* Suppose the lower boundary of the surface layer to be the plane $z = 0$, and the upper to be the plane $z = -T$. The wave is being propagated parallel to the axis of x. Let μ and ρ be the rigidity and density of the upper layer, and let μ' and ρ' refer to the lower layer. Then in both layers u and w are zero, while v is proportional to the real part of $Ve^{\iota\kappa(x-ct)}$, where V may be complex, but is a function of z alone. In both layers δ is zero, and the only equation of motion that is not satisfied identically is

$$-\rho\kappa^2c^2V = \mu\left(-\kappa^2V + \frac{\partial^2 V}{\partial z^2}\right) \qquad \text{.........(1)}$$

for the upper layer, and

$$-\rho'\kappa^2c^2V = \mu'\left(-\kappa^2V + \frac{\partial^2 V}{\partial z^2}\right) \qquad \text{.........(2)}$$

for the lower layer. Putting, as before, $\mu/\rho = \beta^2$, $\mu'/\rho' = \beta'^2$, we have in the lower layer

$$V = Ce^{-s'z} \qquad \text{.........(3),}$$

where C is a constant and

$$\frac{s'^2}{\kappa^2} = 1 - \frac{c^2}{\beta'^2} \qquad \text{.........(4).}$$

In the upper layer

$$V = A\cos sz + B\sin sz \qquad \text{.........(5),}$$

where A and B are constants and

$$\frac{s^2}{\kappa^2} = \frac{c^2}{\beta^2} - 1 \qquad \text{.........(6).}$$

The boundary conditions are that the displacement and the tangential stress are continuous across the lower boundary, and that there is no stress across the free surface. These give

$$C = A \qquad \text{.........(7),}$$

$$-\mu's'C = \mu sB \qquad \text{.........(8),}$$

$$A\sin sT + B\cos sT = 0 \qquad \text{.........(9).}$$

Hence by elimination of A, B, C we have

$$\tan sT = \mu' s'/\mu s \qquad \text{.........(10).}$$

Finally, by substituting for s and s' from (4) and (6), we have the following equation for the wave-velocity:

$$\tan\left\{\kappa T\left(\frac{c^2}{\beta^2}-1\right)^{\frac{1}{2}}\right\} = \frac{\mu'}{\mu}\left(\frac{1-c^2/\beta'^2}{c^2/\beta^2-1}\right)^{\frac{1}{2}} \qquad \text{.........(11).}$$

The condition that the motion shall be indefinitely small at a great depth requires that the real part of s' shall be positive. Hence, by (4), c must be less than β'. Again, if c were less than β, s would be a pure imaginary, by (6), and therefore $s \tan sT$ would be negative. This, by (10), would imply that s' was negative, which is impossible. It follows that c is always greater than β and less than β'. There can therefore be no waves of this type if β is greater than β', for it would then be impossible to find a value of c that would satisfy both of these inequalities. Thus these waves can exist only if the velocity of distortional waves in the surface layer is less than in the matter below.

An optical analogy will illustrate the reason for this result. A wave moving in a surface layer may be compared to a light wave moving nearly parallel to the faces of a plate, with matter of a different refractive index in optical contact with the lower face and a perfect reflecting surface over the upper face. If the velocity of light in the plate is less than that in the underlying matter, light with an angle of incidence on the boundary over the critical angle will be totally reflected, and will continue to be propagated in the plate. If, however, the velocity of light is greater in the plate, light can pass freely from the plate to the underlying matter, whatever its angle of incidence may be. Thus the light will gradually be lost into the underlying material as it advances, and the propagation of a train of waves of constant form will be impossible. In fact Love waves are equivalent to *SH* waves reflected up and down within the upper layer, with associated movement in the lower layer of the type that always arises in total reflexion.

Equation (10) can be written

$$\sigma \tan \sigma\kappa T = \mu'\sigma'/\mu \qquad \text{.........(12),}$$

where σ and σ' have been written for $(c^2/\beta^2-1)^{\frac{1}{2}}$ and $(1-c^2/\beta'^2)^{\frac{1}{2}}$ respectively. When c is equal to β, σ is zero, while σ' is finite and positive. Thus whatever κT may be the left side of (12) will be zero and the right positive. As c increases, the left side will increase steadily until $\sigma\kappa T$ becomes equal to $\frac{1}{2}\pi$, when the left side becomes infinite. If then $(\beta'^2/\beta^2-1)^{\frac{1}{2}}\kappa T$ is greater than $\frac{1}{2}\pi$, the left side will become infinite and positive for some value of c between β and β'. It will then become negative and again increase steadily; if $(\beta'^2/\beta^2-1)^{\frac{1}{2}}\kappa T$ is greater than π but less than $\frac{3}{2}\pi$, the left side will again be positive when c is equal to β'; the discussion can evidently be extended indefinitely. The right side meanwhile decreases steadily from $\mu'(1-c^2/\beta'^2)^{\frac{1}{2}}/\mu$ to zero. Thus the left side

will exceed the right when c is equal to β', unless it has become negative at some intermediate point by passing through an infinite value and has not become positive again by passing through zero; in either case there will be some intermediate value where the left side exceeds the right, and therefore some other value of c that makes the two sides equal. The condition for this is that $(\beta'^2/\beta^2 - 1)^{\frac{1}{2}}\kappa T$ shall be less than π. Similarly we see that there will be two possible values of c that satisfy (12) if $(\beta'^2/\beta^2 - 1)^{\frac{1}{2}}\kappa T$ lies between π and 2π, three if this quantity lies between 2π and 3π, and so on. Again, we see readily that increasing κ while keeping T, μ, and μ' the same makes all the roots approach β. Thus the shorter the wave-length, the more closely will the velocity approach that of distortional waves in the upper layer; and when it becomes indefinitely short these velocities tend to equality. Again, if κ is indefinitely small, we see that the equation (12) can be satisfied only when c approximates to β'; thus waves of great length will travel with velocities approximating to that of distortional waves in the lower layer.

When more than one value of c corresponds to the same value of κ, sz changes by more than π as z changes from 0 to $-T$; hence V, being expressed as a harmonic function of sz with real coefficients, must vanish for some intermediate value of z. Thus there will be no, one, two, or more nodal surfaces within the upper layer according as the root considered is the lowest, second, third or higher value of c corresponding to the actual wave-length.

We see from (7), (8), and (9) that the ratios of A, B, and C are all real; it follows that the motion is in the same phase at all depths.

The type of waves just discussed will be referred to as Love waves. It has been seen that the displacement in them is wholly horizontal and at right angles to the direction of propagation; they should not affect the vertical component. This is in agreement with the appearance of the records. A seismogram of the vertical movement shows a smooth oscillation, which is readily attributable to a single type of wave, namely the Rayleigh waves; but the two horizontal components show violent irregularities, indicating the passage of two sets of waves of widely different periods. It seems probable that the slower of the oscillations shown by the horizontal components are due to Rayleigh waves and to those Love waves that have no nodal plane, while the rapid ones are due to the Love waves with nodal planes.

6·62. The velocity c in 6·61 is the velocity of advance of an unlimited train of harmonic waves all of the same length. An arbitrary initial disturbance may however be regarded as the sum of an infinite number of such trains, each of which then proceeds to spread out with its own proper velocity. If in fact x is a position coordinate and the initial disturbance is

$$u = \frac{1}{\pi}\int_0^\infty \phi(\kappa)\cos\kappa x\, d\kappa \qquad \text{.........(1)}$$

where $\phi(\kappa)$ is a function of κ to be determined by the initial conditions, the disturbance at a later time t is

$$u = \frac{1}{\pi}\int_0^\infty \phi(\kappa) \cos \kappa x \cos \kappa ct \, d\kappa \qquad \text{.........(2)},$$

or, if we put

$$\kappa c = \gamma \qquad \text{.........(3)},$$

$$u = \frac{1}{2\pi}\int_0^\infty \phi(\kappa) \cos(\kappa x - \gamma t) \, d\kappa + \frac{1}{2\pi}\int_0^\infty \phi(\kappa) \cos(\kappa x + \gamma t) \, d\kappa \quad (4).$$

It can be shown that in the class of cases that concern us here the first integral represents a train of waves advancing from the region disturbed initially towards the direction of x increasing, while the second represents a similar train travelling in the opposite direction. An approximation to the first integral* is

$$u = \left\{\frac{1}{2\pi t \,|\, d^2p/d\kappa^2 \,|_0}\right\}^{\frac{1}{2}} \phi(\kappa_0) \cos(\kappa_0 x - \gamma_0 t \mp \tfrac{1}{4}\pi) \qquad \text{.........(5)},$$

where the suffix zero indicates that the corresponding quantity is evaluated for the value of κ that satisfies

$$\frac{x}{t} = \frac{d\gamma}{d\kappa} \qquad \text{.........(6)}.$$

Then $2\pi/\kappa_0$ and $2\pi/\gamma_0$ are the wave-length and period of the waves near the point x at time t. The ratio γ_0/κ_0 is c_0, the wave-velocity of the individual waves in this neighbourhood. Also $(d\gamma/d\kappa)_0$ is the ratio of the distance travelled by the period in question to the time taken, and can therefore be regarded as the velocity of the period. It is not in general equal to c; a given wave usually changes its length and period as it travels. We call $d\gamma/d\kappa$ the *group-velocity* and denote it by C; it is connected with c, the wave-velocity, by the relation

$$C = \frac{d\gamma}{d\kappa} = \frac{d}{d\kappa}(\kappa c) = c + \kappa \frac{dc}{d\kappa} \qquad \text{.........(7)}.$$

6·63. A valuable set of observations of the velocities of Love waves was collected and discussed by Gutenberg†. As what he determined were the epicentral distances and times of the waves of given periods, the velocity determined as their ratio was the group-velocity C; Gutenberg unfortunately interpreted it as c, the wave-velocity, and used the formula 6·61 (11) to determine the unknown T. For Eurasia this was found to be 55 km. The error was pointed out by R. Stoneley‡, who showed that in a relevant case the wave-velocity might be 3·8 km./sec. when the group-velocity was only 3·14 km./sec.; the whole range of variation of the wave-velocity was 3·2 to 4·4 km./sec. The difference between the two velocities was therefore of great importance. The present writer revised the computation§ and found that the thickness determined was reduced to about

* Cf. Havelock, *The Propagation of Disturbances in Dispersive Media*, Cambridge Math. Tracts, 17.

† *Phys. Zs.* **25**, 1924, 377–381; *Der Aufbau der Erde*, 1925, 97.

‡ *M.N.R.A.S. Geoph. Suppl.* **1**, 1925, 280–282. § *Ibid.* 282–292.

15 km. Meanwhile Gutenberg had independently discovered the error*. In later work by Stoneley and Tillotson† the behaviour of Love waves in a crust with two layers of thickness T and T' on top was investigated. With $T = T'$, they found that Gutenberg's data would fit a thickness of 13 km. for each layer; the properties assumed for the intermediate layer, however, corresponded to gabbro rather than to the actual material. A further paper by Stoneley alone‡ assumes, on the basis of the work on near earthquakes, that the intermediate layer is twice as thick as the upper one, and finds 13 km. for the thickness of the upper layer and 26 km. for the lower. The results are reasonably consistent with those given by near earthquakes.

6·64. But as fast as one difficulty in the theory of surface waves is removed, other two take its place. The only hopeful feature is that the new ones generally seem less fundamental than the old one. Looking at the ordinary formula for dispersion 6·62 (5) we see that the coefficient of the cosine is a function of κ_0 only, multiplied by $t^{-\frac{1}{2}}$. Hence for waves of a definite period the amplitude varies as they travel out like $t^{-\frac{1}{2}}$, and therefore like $x^{-\frac{1}{2}}$. This factor is the same for all periods. The amplitudes associated with given periods should therefore be in the same ratio whatever the epicentral distance. Further, the constancy of the group-velocity for a given period is an essential part of ordinary dispersion.

One conspicuous, and hardly surprising, departure from these predictions is that the surface waves show marked damping, whose amount depends on the period. On an average the distance they travel while the damping factor decreases in the ratio $e : 1$ is about 4000–6000 km. But the effect is more severe on the shorter waves than on the longer ones, so that whereas the period of the largest oscillations at short epicentral distances may be about 8 s., that at great distances ($\Delta = 90°$) is on an average 20 s. But the effect is irregular and not even invariably in the same direction§. Some departure from the formula 6·62 (5) is to be expected from the fact that both Love and Rayleigh waves possess a minimum group-velocity. The amplitudes associated with this would be expected to become more prominent with increasing epicentral distance, but the relevant period is about 10 s., so that the effect is in the wrong direction. Frictional damping hardly seems likely to vary as markedly as the data indicate. I have suggested that reflexion and scattering when the waves cross regions where the constitution of the crust changes may produce the requisite effect. There is probably also refraction at the boundaries of continents, which should give great irregularities.

* *Phys. Zs.* **27**, 1926, 111–114.

† *M.N.R.A.S. Geoph. Suppl.* **1**, 1928, 521–527.

‡ *Ibid.* 527–532.

§ J. B. Macelwane, *Bull. Seism. Soc. Amer.* **13**, 1923, 13–69; K. Wegener, *Gött. Nach.* 1912, 333 338.

6·7. The seismological investigation of the sedimentary layer is in its infancy. It does not appear to transmit distinct waves from near earthquakes, but probably plays a part in determining the periods of the pulses and the length of the long train of gradually dying waves that follows the stage of the greatest displacements. Much observational work has been done on the propagation of waves from small explosions and other artificial disturbances, but hitherto it has not been coordinated. Among the chief names associated with it are those of Mintrop and Hubert in Germany, and Maurain, Eblé and Labrouste in France. The stratification of the sedimentary layer probably makes the equations 6·3 (1) incorrect, because they refer to an isotropic solid. Its irregular structure, also, must give rise to much internal reflexion.

6·8. The discussion of the structure of the upper layers up to this point concerns only continental conditions; the evidence of near earthquakes and of the propagation of surface waves in Eurasia specifically concerns the continents. There is abundant geological evidence that the rocks below the ocean floor are systematically different from those of the continents. This is shown especially by the predominance of basic rocks and the rarity of granite on oceanic islands. We shall see later that other evidence is provided by isostasy. It is also probable that sub-oceanic rocks have on the whole cooled more than continental ones. On both grounds we might expect the velocities of seismic waves to be rather higher than in the same material below the continents, and the granitic layer may be thin or absent.

It was found by Angenheister† that the times of transmission of P and S from a Tonga earthquake to a distance of 6° were less by 13 s. and 25 s. respectively than to a similar distance in Europe. The corresponding difference in velocities would be about 15 per cent. This difference, if genuine, does not persist to greater distances; the times of transmission of the waves from the Montana earthquake to distances of 30° under land and under sea were not systematically different. Waves in upper layers analogous to P_g or P^* have not yet been identified under the ocean. If the Fiji Islands, Tonga, and Samoa contained six observatories trustworthy to a second the study of near earthquakes under the ocean might develop, but at present the observatories are too sparsely distributed.

The other source of information about sub-oceanic seismology is provided by the surface waves. The dispersion of Love waves is strikingly different below Eurasia, the Atlantic, and the Pacific. In Eurasia the group-velocities range from about 3·0 to about 4·4 km./sec.; the minimum group-velocity is theoretically slightly less than the velocity of S_g, and the observed velocities cluster strongly about 3·0 km./sec. Under the Pacific group-velocities of 3·8 km./sec. or more are typical for waves with

† *Gött. Nach.* 1921, 1–34.

periods of 20 s., which have average group-velocities of 3·2 km./sec. in Eurasia†, while velocities under 3·5 km./sec. are rare for any period. These facts seem to indicate that the granitic layer is thin or absent under the Pacific. The dispersion under the Atlantic is intermediate in character.

In a recent paper W. Hiller‡ has shown that the maximum amplitude of the Rayleigh waves is propagated 27–28 per cent. faster along oceanic than along continental paths. He remarks that in most cases the relevant period was about 18–20 s. Definite interpretation of his results must await a theoretical discussion of the dispersion of Rayleigh waves.

6·9. *Summary.* On geological grounds it is to be expected that the rocks of the crust increase in basicity with depth, the uppermost igneous layer under the continents being granitic. The study of near earthquakes in continental regions shows that three layers are concerned, the velocities of propagation of compressional and distortional waves in them being as follows: Upper layer, P_g, 5·4–5·6 km./sec.; S_g, 3·3 km./sec.; intermediate layer, P^*, 6·2–6·3 km./sec.; S^*, 3·7 km./sec.; lower layer, P, 7·8 km./sec.; S, 4·35 km./sec. Comparison of the velocities with laboratory determinations of the compressibilities of rocks shows that the upper layer is probably granite, the intermediate one tachylyte (glassy basalt) and the lowest dunite, though other interpretations may be found to fit the seismological data. There seems to be no layer of crystalline basalt. The thicknesses of the upper and intermediate layers are found to be about 10 and 20 km. respectively. The study of surface waves leads to similar results. Conditions below the oceans have been less thoroughly studied, but the evidence indicates that the granitic layer there is thin or absent.

† Cf. the diagram given in several of Gutenberg's recent publications.
‡ *Gerlands Beitr. z. Geophysik*, **17**, 1927, 279–324.

CHAPTER VII

Seismology: The Earth's Interior

"One, two! One, two! And through and through
The vorpal blade went snicker-snack!"
LEWIS CARROLL, *Through the Looking-Glass.*

7·1. *The Zöppritz-Turner Tables.* After Oldham's identification of the three phases in records of distant earthquakes, and his construction of preliminary time-curves for them, the next great advance on the observational side was by K. Zöppritz*, who collected the available information for three well-observed earthquakes and put it together so as to give a new set of time-curves. The earthquakes used were the Indian one of 1905 April 4, the Calabrian one of 1905 September 8, and the Californian one of 1905 April 18. The results for epicentral distances under 1000 km. depended wholly on the Calabrian earthquake, and it is worth while to rediscuss them in relation to the modern work on near earthquakes. Zöppritz proceeded by assuming the velocity of P from the epicentre to Messina to be 7 km./sec., and determined the time at the origin accordingly. This time being taken as 1 h. 43 m. 5 s., he obtained the times of transit T in the third column of the following table. The fourth column Δ/T then gives the mean apparent velocities to various distances in km./sec. Their consistency clearly leaves something to be desired. We now know, however,

	Δ (km.)	T	Δ/T	$T - \Delta/7{\cdot}8$
Messina	80	12	6·7	2
Catania	170	25	6·8	3
Ischia	270	42	6·5	7
Rocca di Papa	430	55	7·8	0
Wien	1050	133	7·9	1
München	1100	158	7·0	17
Jena	1390	177	7·9	1

that at so short an epicentral distance as 80 km. the first pulse to arrive would normally be P_g and not P, and P_g was therefore probably the wave recorded at Messina. If we form the differences $T - \Delta/7{\cdot}8$ we obtain the fifth column of the table. The results for Ischia and München are now seen to be definitely exceptional, and the approximate constancy of the others indicates that a velocity of 7·8 km./sec. was very near the truth. We may take the time of starting for P as $T - \Delta/7{\cdot}8$ evaluated for Catania, and for P_g as $T - \Delta/5{\cdot}6$ evaluated for Messina. The latter is -2 s., so that the delay of starting of P in comparison with P_g was 5 s.; this compares well with the 6–7 s. given by the recent studies of near earthquakes.

* *Gött. Nach.* 1907, 529–549.

In Zöppritz's final results a velocity of 7·2 km./sec. is ascribed to P up to $\Delta = 666$ km. At this distance the time of transit according to him is 92 s.; while $\Delta/7{\cdot}8$ is 85 s., so that a difference of 7 s. has accumulated owing to the adoption of too low a velocity at short epicentral distances.

The times of transmission to greater distances were obtained from the Calabrian earthquake, which was observed up to $\Delta = 7660$ km., supplemented by the observations of the other two. The most distant observation was that of the Californian earthquake at Tiflis, 11,110 km. away. There were conspicuous gaps where no observations were available, the chief being from 5680 km. to 7660 km. These were filled up by interpolation, and the curve for P was continued up to 13,000 km. by extrapolation. The same applies to the treatment of S, which however was observed at no epicentral distance under 1185 km. At this distance the observed difference between the times of arrival of P and S was 122 s.; that calculated on the supposition that they started at the same time and had velocities 7·8 and 4·35 km./sec. would be 121 s. Clearly there is no inconsistency between Zöppritz's data and the results obtained from near earthquakes. But the velocity of S finally adopted by him for distances up to 666 km. was only 4·0 km./sec., owing to the fact that the apparent times of starting had been made too early in order to fit the low velocity assumed for P.

Though Zöppritz's curves are affected by these systematic errors they have formed the basis of the chief developments in the study of distant earthquakes since their publication. They were reduced to the form of tables by Prof. H. H. Turner (and incidentally extrapolated further to $\Delta = 150°$)*. In this form they have been used in the reduction of seismological observations in the *British Association Bulletin* and the *International Seismological Summary* since these publications began. Their general correspondence with the facts is good, though it is becoming clear that they require corrections of the order of 20 s. at some epicentral distances. In the meantime several other tables of times of transmission have been published, but Prof. Turner has, wisely in my opinion, continued to use the old ones until it should be possible to feel confident that a change once made was likely to be final. The times at present in use are given in the table on p. 119.

7·2. Epicentral distances are usually calculated from the formula of spherical trigonometry,

$$\cos \Delta = \cos \theta \cos \theta' + \sin \theta \sin \theta' \cos (\phi - \phi') \quad \text{.........(1)},$$

* Zöppritz gave his epicentral distances in kilometres, measured as arcs along the surface of the earth, and this practice is still followed by most continental seismologists. The usual British practice is to give the distance in degrees, as the angle subtended at the centre of the earth by the arc. To convert from one to the other, notice that 9° = 1000 km., or 1° = 111·1 km.

Degrees	*P* sec.	*S* sec.	*S − P* sec.	Degrees	*P* sec.	*S* sec.	*S − P* sec.	Degrees	*P* sec.	*S* sec.	*S − P* sec.
1	15	28	13	51	553	991	438	101	855	1565	710
2	31	55	24	52	560	1004	444	102	860	1575	715
3	47	83	36	53	566	1016	450	103	865	1584	719
4	62	110	48	54	573	1029	456	104	870	1593	723
5	77	137	60	55	579	1041	462	105	874	1602	728
6	92	164	72	56	586	1054	468	106	879	1612	733
7	106	190	84	57	592	1066	474	107	884	1621	737
8	121	217	96	58	599	1079	480	108	888	1630	742
9	136	243	107	59	605	1091	486	109	893	1639	746
10	150	269	119	60	612	1103	491	110	897	1648	751
11	164	294	130	61	619	1116	497	111	902	1657	755
12	179	319	140	62	625	1128	503	112	907	1666	759
13	193	344	151	63	632	1141	509	113	911	1674	763
14	206	368	162	64	638	1153	515	114	916	1682	766
15	219	392	173	65	645	1165	520	115	920	1690	770
16	232	415	183	66	651	1177	526	116	925	1698	773
17	245	438	193	67	658	1190	532	117	929	1706	777
18	257	460	203	68	664	1202	538	118	934	1714	780
19	269	482	213	69	671	1214	543	119	938	1722	784
20	281	503	222	70	677	1226	549	120	942	1729	787
21	293	524	231	71	683	1238	555	121	947	1737	790
22	305	545	240	72	690	1250	560	122	952	1744	792
23	317	565	248	73	696	1262	566	123	957	1752	795
24	328	584	256	74	702	1274	572	124	961	1759	798
25	338	603	265	75	709	1286	577	125	966	1766	800
26	348	622	274	76	715	1297	582	126	970	1773	803
27	358	641	283	77	721	1309	588	127	974	1780	806
28	368	659	291	78	727	1320	593	128	978	1787	809
29	378	677	299	79	733	1332	599	129	983	1794	811
30	388	694	306	80	739	1343	604	130	988	1801	813
31	398	711	313	81	745	1355	610	131	992	1807	815
32	407	728	321	82	750	1366	616	132	996	1814	818
33	416	744	328	83	756	1377	621	133	1001	1821	820
34	425	760	335	84	762	1388	626	134	1005	1827	822
35	433	775	342	85	768	1399	631	135	1009	1833	824
36	442	790	348	86	773	1410	637	136	1014	1840	826
37	450	804	354	87	779	1421	642	137	1018	1846	828
38	458	818	360	88	785	1432	647	138	1023	1852	829
39	466	832	366	89	790	1443	653	139	1027	1858	831
40	475	847	372	90	796	1454	658	140	1031	1864	833
41	483	861	378	91	801	1464	663	141	1035	1869	834
42	491	875	384	92	807	1475	668	142	1039	1875	836
43	498	888	390	93	812	1485	673	143	1043	1881	838
44	506	902	396	94	818	1496	678	144	1047	1886	839
45	513	915	402	95	823	1506	683	145	1051	1892	841
46	520	928	408	96	829	1516	687	146	1055	1897	842
47	527	941	414	97	834	1526	692	147	1059	1902	843
48	534	954	420	98	840	1536	696	148	1063	1907	844
49	540	966	426	99	845	1546	701	149	1067	1912	845
50	547	979	432	100	851	1556	705	150	1071	1917	846

where θ and θ' are the colatitudes, ϕ and ϕ' the east longitudes, of the observing station and the epicentre. If we write

$$a = \sin\theta\cos\phi; \quad b = \sin\theta\sin\phi; \quad c = \cos\theta \quad \text{.........(2)},$$

$$A = \sin\theta'\cos\phi'; \; B = \sin\theta'\sin\phi'; \; C = \cos\theta' \quad \text{.........(3)},$$

(a, b, c) and (A, B, C) are the direction cosines of the radius of the earth drawn to the station and the epicentre, the axes being the equatorial diameters in the Greenwich meridian and 90° east of it, and the polar axis. Then (1) becomes the simple formula of solid geometry

$$\cos\Delta = aA + bB + cC \qquad \text{.........(4)}.$$

Prof. H. H. Turner has put this into a very convenient form for computation* by noticing that it is equivalent to

$$2(1 - \cos\Delta) = (a - A)^2 + (b - B)^2 + (c - C)^2 \quad \text{.........(5)}.$$

The method of use is to have (a, b, c) computed once for all for each observatory, and to calculate (A, B, C) from the assumed latitude and longitude of the epicentre. Then $a - A$, $b - B$, and $c - C$ are found by subtraction and $2(1 - \cos\Delta)$ by the use of a table of squares and addition. A table of $2(1 - \cos\Delta)$ has been provided. The method is much more convenient than the direct use of (1). The constants (a, b, c) for the stations are published to three places of decimals in a supplement to the *International Seismological Summary*, though it would be convenient for work on near earthquakes to have them to four places.

If Δ is over 90° it is often more convenient to use the formula

$$2(1 + \cos\Delta) = (a + A)^2 + (b + B)^2 + (c + C)^2 \quad \text{.........(6)}.$$

7·3. *Velocities in the interior of the earth.* Observations of distant earthquakes can be used in two ways to give information about the velocities of elastic waves in the interior of the earth. One depends wholly on the use of the times of transmission, while the other supplements these considerably by using also the amplitudes of the motions. We consider a wave spreading out from a point on the surface of the earth, and suppose its velocity in the neighbourhood of any internal point to be c, and the time taken to reach that point to be T. T will evidently be a function of the position of the point. Then the surfaces where T is constant will mark the consecutive positions of the wave front. We shall treat the earth as spherically symmetrical, so that c is a function of the distance from the centre alone. Let r denote the distance of a point from the centre, and θ the angle between the lines joining the point and the focus to the centre. Thus r and θ are spherical polar coordinates. The wave front will at all instants be symmetrical about the line joining the centre of the earth to

* *M.N.R.A.S.* **75**, 1915, 530–541.

the focus of the earthquake. The time taken by the wave to reach a given point is

$$T = \int \frac{ds}{c}$$

$$= \int \frac{1}{c} \left\{ \left(\frac{dr}{d\theta} \right)^2 + r^2 \right\}^{\frac{1}{2}} d\theta \qquad \text{.........(1)},$$

where ds is an element of length along the path actually taken by the wave in passing from the focus to the point. The actual path is the one that makes this time the shortest possible, since the records of earthquakes give the time of the commencement of each phase. Thus the actual path has to be such as to make the integral (1) a minimum. Putting for a moment V for the integrand in (1) and ρ for $dr/d\theta$, we know from the calculus of variations that the integral is stationary if r satisfies the differential equation

$$\frac{\partial V}{\partial r} - \frac{d}{d\theta} \frac{\partial V}{\partial \rho} = 0 \qquad \text{.........(2)},$$

a first integral of which is known* to be

$$V = \rho \frac{\partial V}{\partial \rho} + p \qquad \text{.........(3)},$$

where p is a constant. On substituting for V and simplifying we find

$$\frac{r^2}{c} = p \left\{ \left(\frac{dr}{d\theta} \right)^2 + r^2 \right\}^{\frac{1}{2}} \qquad \text{.........(4)},$$

whence

$$\frac{dr}{d\theta} = \pm r \left\{ \frac{r^2}{p^2 c^2} - 1 \right\}^{\frac{1}{2}} \qquad \text{.........(5)},$$

and

$$\theta = \pm \int^r \frac{p\, dr}{r \left\{ \frac{r^2}{c^2} - p^2 \right\}^{\frac{1}{2}}} \qquad \text{.........(6)}.$$

The disturbance commences by moving inwards, so that $dr/d\theta$ is negative; thus the negative sign must be taken for the root. Again, $dr/d\theta$ vanishes when the ray reaches its nearest point to the centre. We see by (5) that the value of r at this point is pc. If χ is the corresponding value of θ, we have

$$\chi = \int_{pc}^{R} \frac{p\, dr}{r \left\{ \frac{r^2}{c^2} - p^2 \right\}^{\frac{1}{2}}} \qquad \text{.........(7)},$$

where R is the radius of the earth. After passing this point the ray bends upwards, remaining symmetrical about the line joining the centre to the point nearest to the centre, and reaches the surface again at the point $(R, 2\chi)$. Evidently 2χ is the angular distance between the focus and the point of emergence of the ray, and may be denoted by Δ. The tables give the time taken by the ray in reaching this point as a function of Δ.

* Todhunter, *Integral Calculus*, 1883, p. 346.

Now let P be the point (R, Δ), and T the time taken in reaching it. Let P' be a neighbouring point $(R, \Delta + d\Delta)$, and $T + dT$ the time taken in reaching it. Δ is an angle, in circular measure. Draw PQ perpendicular to the ray that reaches P', and let c_0 be the velocity of the wave near the surface. Then

$$QP' = c_0 dT; \quad PP' = R d\Delta \qquad \text{.........(8)},$$

and therefore

$$\frac{QP'}{PP'} = \frac{c_0}{R}\frac{dT}{d\Delta} \qquad \text{.........(9)}.$$

This ratio, however, is the cosine of the angle made with the surface by the emergent ray, and therefore, on account of the symmetrical form of the ray, is equal to the cosine of e, the angle made with the surface by the ray when it enters at the focus. The latter is equal to the value of $Rd\theta/ds$ when r is equal to R. But by (4)

$$\frac{d\theta}{ds} = \frac{pc}{r^2} \qquad \text{.........(10)}.$$

Hence we must have

$$\frac{c_0}{R}\frac{dT}{d\Delta} = \frac{pc_0}{R},$$

or

$$p = \frac{dT}{d\Delta} \qquad \text{.........(11)}.$$

Thus p is a calculable function of Δ, and therefore of χ. Therefore χ may be expressed as a function of p, $f(p)$ say. Equation (7) therefore contains only one unknown, namely c. Our problem is therefore to determine c as a function of r from this integral equation. If we put

$$\frac{r}{c} = \eta \qquad \text{.........(12)},$$

equation (7) becomes

$$f(p) = p\int_p^{R/c_0} \frac{\frac{d}{d\eta}(\log r)\, d\eta}{(\eta^2 - p^2)^{\frac{1}{2}}} \qquad \text{.........(13)},$$

which has been solved by Herglotz and Bateman*. The solution is

$$\log r = C - \frac{2}{\pi}\int_\eta^{R/c_0} \frac{f(\mu)\, d\mu}{(\mu^2 - \eta^2)^{\frac{1}{2}}} \qquad \text{.........(14)},$$

where C is a constant. When η approaches R/c_0, r approaches R, and therefore

$$C = \log R.$$

Hence finally

$$\log\frac{R}{r} = \frac{2}{\pi}\int_\eta^{R/c_0} \frac{f(\mu)\, d\mu}{(\mu^2 - \eta^2)^{\frac{1}{2}}} \qquad \text{.........(15)}.$$

This equation gives r as a function of η, while, by (12),

$$c = r/\eta \qquad \text{.........(16)},$$

so that a correspondence can be established between the values of r and c, and c is therefore obtained as a function of r. The calculation is laborious,

* G. Herglotz, *Phys. Zs.* **8**, 1907, 145–147; H. Bateman, *Phil. Mag.* Ser. 6, 1910, 576–587.

but has been applied to the Zöppritz-Turner tables by C. G. Knott*. The numerical integration can be somewhat simplified by a device due to Wiechert†. We notice that in (13) η and r are quantities that vary along a given ray, p remaining a constant determined by the angle of emergence. In (14), on the other hand, η and r refer to the deepest point of the ray, while μ, which replaces p, ranges from η to R/c_0, and therefore gives the angles of emergence of the rays that emerge at distances Δ_1 between 0 and Δ. If we put

$$\mu = \eta \cosh q \qquad \text{.........(17)},$$

we can write $\quad \mu = (dT/d\Delta)_1;\ \eta = (dT/d\Delta);\ f(\mu) = \tfrac{1}{2}\Delta_1 \quad$ ....(18).

Thus

$$\cosh q = \frac{(dT/d\Delta)_1}{(dT/d\Delta)} = \frac{\cos e_1}{\cos e} \qquad \text{.........(19)},$$

and q ranges from 0 to $\cosh^{-1} \sec e$. Thus (15) becomes

$$\log \frac{R}{r} = \frac{1}{\pi} \int_0^{\cosh^{-1} \sec e} \Delta_1 dq$$

$$= \frac{1}{\pi} [\Delta_1 q] - \frac{1}{\pi} \int_\Delta^0 q\, d\Delta_1 \qquad \text{.........(20)}.$$

But q vanishes at one limit and Δ_1 at the other, so that the integrated part is 0 and

$$\log \frac{R}{r} = \frac{1}{\pi} \int_0^\Delta q\, d\Delta_1 \qquad \text{.........(21)}.$$

Thus when q is found by (19) for values of Δ_1 less than Δ, the depth reached by the ray emerging at distance Δ is given directly by (21). Also by (16) the value of r/c at this depth is $dT/d\Delta$ and therefore

$$c = r \Big/ \frac{dT}{d\Delta} \qquad \text{.........(22)}.$$

Wiechert's method was applied by S. Mohorovičić to the revised tables for P constructed by his father and himself‡.

7·31. The use of amplitudes to improve the time curves was begun by Zöppritz, and carried out by L. Geiger and Gutenberg§. The method is based on the assumptions that the focus sends out energy equally in all directions, and that the propagation of energy is wholly along the rays and not across them. Let us suppose that the energy emitted per unit solid angle is K. We use the angle of emergence of the waves in preference to the angle of incidence because it vanishes with Δ. The angle of incidence of a ray on entering is, by symmetry, the same as on emergence. The energy sent out within a range de is $2\pi K \cos e\, de$. The area of the surface

* *Proc. Roy. Soc. Edin.* **39**, 1919, 158–208.

† Wiechert and Geiger, *Phys. Zs.* **11**, 1910, 294–312.

‡ *Gerlands Beitr.*, 13, 1914, 217–240 and 14, 1916, 187–198.

§ *Gött. Nach.* 1912, 121–206, 623–675.

of the earth between epicentral distances Δ and $\Delta + d\Delta$ is $2\pi R^2 \sin \Delta\, d\Delta$. The energy received per unit surface is therefore

$$\frac{K}{R^2}\frac{\cos e}{\sin \Delta}\frac{de}{d\Delta} \qquad \text{.........(1).}$$

But the angle between the wave front and the surface is $\frac{1}{2}\pi - e$, so that unit surface corresponds to an area $\sin e$ of the wave front. Hence the energy per unit area of the wave front is

$$\frac{K}{R^2}\frac{\cot e}{\sin \Delta}\frac{de}{d\Delta} \qquad \text{.........(2).}$$

If a shock of the same intensity occurred in a homogeneous body of infinite extent, the energy per unit area of the wave front at distance R from the focus would be K/R^2. Taking this as our standard of intensity, we have for the intensity of the wave coming up to the surface the formula

$$E = \frac{\cot e}{\sin \Delta}\frac{de}{d\Delta} \qquad \text{.........(3).}$$

Also e satisfies the relation $\quad \cos e = \dfrac{c_0}{R}\dfrac{dT}{d\Delta} \qquad$ ....(4).

The amplitude of the wave coming up to the surface is proportional to $\sqrt{E}$. To obtain the motion of the ground we must apply a factor to allow for the ratio of the movement of the ground to that in the emergent wave. It may also be necessary to allow for losses of energy by absorption and reflexion on the way.

For an origin at the surface $0 < e < \frac{1}{2}\pi$. Also E is necessarily positive, and therefore by (3) $de/d\Delta$ is positive. Differentiating (4) we therefore see that $d^2T/d\Delta^2$ is negative. So long as the laws of geometrical optics apply the time curves are concave to the axis of Δ at all places where waves are received.

If the amount of displacement of the ground was accurately known as a function of E and e, and was also accurately observed, (3) would be a differential equation giving e in terms of Δ, soluble by direct integration. Then a further integration of (4) gives T. We can either obtain the times of transmission by two integrations based on the observed amplitudes, or the amplitudes by two differentiations based on the observed times. But the quantity that interests us most directly, on account of the information it gives on the earth's internal structure, is e; for by 7·3 (9)

$$\cos e = \frac{c_0 dT}{R d\Delta} = \frac{c_0}{R}\frac{r}{c} \qquad \text{.........(5),}$$

and is therefore directly connected with the properties of the earth at the deepest point reached by the ray. To investigate the structure of the earth, therefore, we can either integrate the amplitude curve or differentiate the time curve. The latter gives the method of Herglotz and Bateman, the former that of Zöppritz, developed by Geiger and especially Gutenberg.

tation of the variation in the times of transit of [P] clearly requires further examination.

7·8. The velocities of P and S at various depths given in Gutenberg's *Lehrbuch* (1927) are shown by the continuous lines in Fig. 10, while the dotted lines show the results of C. G. Knott based on the Zöppritz-Turner tables. The results for P obtained by S. Mohorovičić in 1916 are not shown because he has accepted his father's new time-curves, which are almost indistinguishable from those of Gutenberg. With regard to Knott's results, it must be said that the part of his curve for P below a depth of 2700 km., and that for S below 2800 km., depend on the extrapolated part of the Zöppritz curves. Within the region where data exist there is a general correspondence between the results, though the velocities found by Knott are on the whole too low. Adoption of the revised times given by myself would make the P curve start from 7·8 instead of 8·0 km./sec.; the velocities of both P and S about 1200 km. down would be slightly increased, and those between 2400 km. and 2900 km. would be slightly reduced.

Adams and Gibson estimate that the change of compressibility of dunite per 10^9 c.g.s. units of pressure is 0·007 of the compressibility at 2000 megabars. This would make the velocity of elastic waves increase with depth at the rate of 1 part in 850 per km. The actual increase for both P and S is about one-half in the first 1200 km., so that the variation indicated by work in the laboratory is not maintained. At greater depths the increase is of course slower still.

7·9. *Summary*. The observations of P and S at short distances used in the construction of the Zöppritz-Turner tables have been seen to be consistent with the views expressed here concerning the structure of the upper crust. The methods of applying the variations of the times of transmission and the amplitudes of the displacement to give information about the velocities of P and S waves in the interior of the earth have been described. The most striking feature revealed is the presence of a central core, apparently liquid, with a radius rather more than half that of the earth as a whole, and with a sharp boundary. The derivation of reflected and transformed waves, and the suggested improvements in the adopted times of transmission of the waves, are described.

CHAPTER VIII

The Thermal History of the Earth

"I know it's something humorous, but lingering."
W. S. GILBERT, *The Mikado*.

8·1. We have seen that the radiation from the surface of the earth when in a liquid state must have been intense enough to dispose of its internal heat, including the latent heat of solidification, in a few thousand years. The method of solidification is of some importance. So long as the whole was fluid the process was comparatively simple. The heavy materials of the iron core quickly settled to the centre and stayed there. Liquids in general contract and become denser as they cool, and the matter cooled at the surface by radiation quickly sank through the hotter liquid below, thus maintaining irregular convection currents and a continual supply of heat to the surface. It is probable, though not certain, that stirring by the currents would keep the whole of the rocky shell uniform in composition. The distribution of temperature can be estimated from thermodynamical considerations. If we consider unit mass of the substance, the inflow of heat satisfies the conditions

$$dQ = dE + pdv = d(E + pv) - vdp = c_p dV + Mdp \quad \text{.........(1)},$$

where E is the internal energy, v the volume, p the pressure, V the absolute temperature, c_p the specific heat at constant pressure, and M is to be found. The condition that $d(E + pv)$ shall be a perfect differential gives

$$\frac{\partial}{\partial V}(M + v) = \frac{\partial c_p}{\partial p} \quad \text{.........(2)},$$

and the existence of $\int dQ/V$, the entropy, gives

$$\frac{\partial}{\partial p}\left(\frac{c_p}{V}\right) = \frac{\partial}{\partial V}\left(\frac{M}{V}\right) \quad \text{.........(3)}.$$

Combining (2) and (3) we find

$$M = -V\frac{\partial v}{\partial V} \quad \text{.........(4)},$$

where v is regarded as a function of the pressure and temperature, so that $\partial v/\partial V$ is α/ρ, where α is the coefficient of thermal expansion by volume, and ρ the density. Now if no heat is gained by conduction (1) gives

$$\frac{dV}{dp} = -\frac{M}{c_p} = \frac{\alpha V}{\rho c_p} \quad \text{.........(5)}.$$

This gives the rate of increase of temperature on adiabatic compression*. In a fluid cooling by convection this relation is satisfied very accurately. If, further, x is the depth,

$$dp/dx = g\rho \qquad \text{.........(6)}$$

and

$$\frac{dV}{dx} = \frac{g\alpha V}{c_p} \qquad \text{.........(7).}$$

With $g = 981$ cm./sec.^{2}, $\alpha = 2 \times 10^{-5}$ per degree, $V = 1400°$, $c_p = 0{\cdot}2$ cal./gm. degree $= 8 \times 10^6$ ergs/gm. degree, this gives $dV/dx = 3 \times 10^{-6}$ degree per centimetre or 0·3 degree per kilometre. The estimate is rough, but it will serve our purpose.

During the liquid stage the vertical variation of temperature according to (7) would be maintained. If the gradient became greater, convection currents would increase in vigour and redistribute the temperature adiabatically; if it became less, convection currents would be damped down by viscosity, cooling would be confined to the top, and the gradient would steepen. The whole of the liquid mass would therefore cool together.

8·11. The melting point of the material would depend on the pressure and therefore on the depth. The effect of pressure on the melting point is expressed by the equation

$$\frac{dV_0}{dp} = \frac{V_0}{L}\left(\frac{1}{\rho_1} - \frac{1}{\rho_2}\right) \qquad \text{.........(1),}$$

where V_0 is the melting point, L the latent heat of fusion, and ρ_1 and ρ_2 the densities of the substance in the liquid and solid states. Thus in a liquid layer

$$\frac{dV_0}{dx} = \frac{gV_0}{L}\left(1 - \frac{\rho_1}{\rho_2}\right) \qquad \text{.........(2).}$$

With the fairly typical values for silicate rocks $V_0 = 1300°$, $L = 100$ cal./gm. $= 4 \times 10^9$ ergs/gm., $\rho_1/\rho_2 = 0{\cdot}9$, this gives 3×10^{-5} degrees C. per centimetre or 3° per kilometre.

8·12. The important feature of 8·1 (7) and 8·11 (2) is that in a well-stirred fluid magma, while the actual temperature and the melting point both increase with depth, the melting point increases the faster. As the fluid cools, therefore, the melting point is first reached at the bottom, and solidification proceeds from the bottom upwards. Convective agitation in the fluid continues until solidification is complete; the rate of solidification is determined by the condition that the heat given out by the liquid at the bottom in solidifying and by the body of the liquid in cooling must supply the actual loss by radiation from the outside.

* L. H. Adams, *J. Wash. Acad. Sci.* **14**, 1924, 459–472. The factor ρ is omitted from Adams's formula.

The iron core hardly affects these considerations. There would always be continuity of temperature across its boundary, but the difference of density is too great for convective interchange between core and shell, though convection currents would exist within both and would maintain conditions of the type 8·1 (7) or 8·11 (2) in each. The core and the shell would cool together until solidification began and stopped convection at the bottom of the shell. Thenceforward the core could cool only by conduction through the solid, and further loss of heat from it became unimportant.

8·13. The inference that solidification proceeded from the bottom upwards was made by L. H. Adams, in the paper just cited. He called attention also to another consideration of importance in some connexions*, though not, I think, in this. The equation 8·1 (7) holds strictly only for a fluid agitated in such a way that no heat at all is interchanged between elements. In an actual fluid cooling by convection there is interchange of heat by conduction; also the viscosity of the fluid tends to damp out the agitation, and some means is required to maintain it against conduction and viscosity. Adams suggests that a viscosity of the order of 10^7 c.g.s. would suffice to prevent convective agitation, which would therefore cease when the liquid reached this viscosity, probably at a temperature appreciably above the melting-point. Cooling by conduction would replace convection at this stage; thus a layer still liquid, but too viscous to cool by convection, would surround the iron core, and would thicken with time at the expense of the mobile layer above.

The viscosity needed to prevent convective agitation can be estimated. In a layer of depth H, with vertical temperature gradient β, thermometric conductivity κ, and kinematic viscosity ν, convection begins when

$$\frac{g\alpha\beta H^4}{\kappa\nu} = \lambda,$$

where λ is a number depending on the boundary conditions, but usually about 1000†. When the effect of pressure on the temperature is allowed for, β must be replaced by the excess of the actual gradient over the adiabatic. With $g = 1000$, $\alpha = 2 \times 10^{-5}$, $\kappa = 0{\cdot}01$, $H = 2{\cdot}5 \times 10^8$, $\beta = 10^{-5}$, in c.g.s. Centigrade units, ν would have to be of order 10^{26} cm.^{2}/sec. to prevent convective agitation. This is a viscosity far greater than can be measured, and it appears that in a layer as deep as the rocky shell, with a temperature gradient appreciably above the adiabatic, convective agitation could be prevented only by a viscosity so great that the substance would be indistinguishable from a solid. Such a viscosity would presumably be attained at a temperature very close to the melting point, and

* See later, 145–148.

† Rayleigh, *Phil. Mag.* **32**, 1916, 529–546; Jeffreys, *Phil. Mag.* **2**, 1926, 833–844; *Proc. Roy. Soc.* A, **118**, 1928, 195–208.

therefore even if Adams's suggestion is correct (as it probably is) convection would cease at a temperature gradient very near that given by 8·11 (2).

8·14. The above discussion supposes that the material of the shell was originally thoroughly mixed, and remained so throughout solidification, that the gradients given by 8·1 (7) and 8·11 (2) were constant, and that there was no internal generation of heat. The first postulate is probably true; most of the others are certainly false. The variability with depth of the quantities in 8·1 (7) is probably not serious. But the factor $1 - \rho_1/\rho_2$ in 8·11 (2) probably varies considerably. A pressure of $1{\cdot}2 \times 10^{11}$ dynes/cm.^{2}, which would be reached at a depth of 400 km., would increase the density of solid dunite by about 10 per cent. The compressibility in the liquid state is presumably greater, perhaps double, and the difference in density at atmospheric pressure is unlikely to exceed 10 per cent. Therefore if the difference in compressibilities is maintained to high pressures the difference in densities in the solid and liquid states will be annihilated at a depth of a few hundred kilometres, and the effect of pressure on the melting point will disappear also. It is more likely that the compressibilities become nearly equal, and that the solid is always denser than the liquid, but in any case dV_0/dx at great depths is much less than at the surface. If, as is quite likely, it is less than the adiabatic gradient in the liquid, the cooling liquid will reach the melting point first at an intermediate depth, probably of some hundreds of kilometres, and will begin to solidify there. Above this level solidification will proceed as if from the bottom. But below it the escape of heat can only occur by conduction through a steadily thickening solid layer, and a thick liquid layer will be preserved for a long time.

The separation of the granitic and intermediate layers of the crust must be taken into account; so must radioactivity in the rocks. After the upper layers separated convection could proceed only within each layer separately; currents rising all the way from the boundary of the core to the outer surface would no longer be possible. In the more constricted conditions of the upper layers the reduced value of H would cause convection to stop at a much lower viscosity; for $H = 30$ km. and a difference of 30° between the top and bottom, the critical viscosity is about 10^{18} cm.^{2}/sec.* But as the consolidation points of olivine and granite differ by some hundreds of degrees it seems that the granite would remain entirely liquid till after the lower layer had begun to solidify.

8·2. Radioactivity is a more serious consideration. It has been explained already that radioactive elements in their degeneration liberate helium. This comes off in the form, not of neutral helium atoms, but of

* For comparison we may remark that the viscosity of water is about 10^{-2}; olive oil at 15° C., 1; Lyle's golden syrup, 10^3; shoemaker's wax, 10^6; pitch at 15°, 10^{10}; plate glass, over 10^{18} in the same units.

α particles, which are helium atoms without their two outer electrons. The velocities of emission of these particles have been measured. It is found that all α particles emitted by the same element have the same velocity. The number emitted per second per gram of radium can be found by direct counting, and the mass of an α particle is known. Hence the total mass and the velocity of the particles emitted are known, and therefore their kinetic energy can be found. Again, it has already been explained that in any actual rock the quantities of all the radioactive substances present are such that the same number of atoms of each break up every second. Hence if we know the amount of any one radioactive element in a rock we can find the rate of production of α particles by each member of the series. The velocities of the α particles from all members being known, the total kinetic energy of those from each element can be found, and hence finally if the quantity of any one radioactive element of the uranium series in a rock is known, the kinetic energy of all the α particles emitted per second by all the members of the uranium series present can be calculated.

8·21. Now in an actual rock an emitted α particle cannot proceed far before it is stopped by the surrounding material. Its kinetic energy then becomes converted into heat. Thus the presence of a known quantity of radioactive materials in a rock enables us to infer the rate of supply of heat to that rock. Indeed, one of the first facts noticed about radium was that its temperature was always a trifle above that of its surroundings. The result is that for every gram of uranium present, uranium and its products produce $8{\cdot}0 \times 10^{-5}$ calories per hour. This result requires to be increased somewhat to allow for the fact that α particles are not the only form of radiation from radioactive substances. In addition they send out β particles, which are free electrons, and γ rays, which are electromagnetic waves closely resembling X-rays. The energy liberated by the absorption of these in the medium is enough to increase the estimate just mentioned to $8{\cdot}5 \times 10^{-5}$ calories per gram per hour. An important check on this was provided by H. H. Poole, who filled a Dewar flask with pitchblende, kept the outside at a fixed temperature, and determined the ultimate steady difference of temperature between the inside and the outside by means of a thermocouple. He found the total rate of emission to be 10^{-4} cal./gm.hr. The measurement was, however, one of extreme difficulty, because the difference of temperature involved was only $0^\circ{\cdot}007$. An additional check on a part of the work has been provided by the measurements of Rutherford and Robinson, and St Meyer and Hess, of the heat emitted by radium and its degeneration products. This agrees with the value calculated from the properties of the emitted rays within at most 1 per cent.* It therefore seems that the calculated generation of heat by radioactive elements must be very accurate.

* Lawson, *Nature*, **119**, 1927, 277 and 703.

The rates of generation of heat by uranium and thorium, with their degeneration products, are respectively $2{\cdot}5 \times 10^{-8}$ and $0{\cdot}72 \times 10^{-8}$ cal./gm.sec. Potassium also is radioactive, but very feebly so. Its decay constant is certainly large compared with the age of the earth, and it produces* about $3{\cdot}9 \times 10^{-12}$ cal./gm. sec. But it is so abundant in comparison with uranium and thorium that in aggregate output of heat it is comparable with them. The actual amounts of radioactive elements present in rocks, even of the same type, are variable. Thus in two series of granites, 22 specimens in all, examined by J. H. J. Poole and J. Joly†, the largest radium content was six times the lowest, and the largest thorium content twelve times the lowest. Nevertheless representative values can be found with reasonable care. The following values are partly from a discussion by Holmes and Lawson‡, and partly directly and partly indirectly from the above work of Poole and Joly. Data for potassium in diorites, basalts, and dunite are from Clarke's *Geochemistry*. The value for potassium in basaltic rocks is a mean for eight gabbros. Effusive basalts may have much more potassium. Poole and Joly used only two specimens of dunite, which were, however, very similar.

Rock type		U	Th	K	Total
Granite	Content	$9{\cdot}0 \times 10^{-6}$	$20{\cdot}0 \times 10^{-6}$	$3{\cdot}4 \times 10^{-2}$	
	Heat generation	$22{\cdot}0 \times 10^{-14}$	$14{\cdot}0 \times 10^{-14}$	$13{\cdot}0 \times 10^{-14}$	49×10^{-14}
Diorite, andesite, etc.	Content	$6{\cdot}6 \times 10^{-6}$	$6{\cdot}1 \times 10^{-6}$	$2{\cdot}2 \times 10^{-2}$	
	Heat generation	$16{\cdot}5 \times 10^{-14}$	$4{\cdot}4 \times 10^{-14}$	$8{\cdot}5 \times 10^{-14}$	29×10^{-14}
Basalt, dolerite, gabbro	Content	$3{\cdot}5 \times 10^{-6}$	$7{\cdot}7 \times 10^{-6}$	$0{\cdot}76 \times 10^{-2}$	
	Heat generation	$8{\cdot}7 \times 10^{-14}$	$5{\cdot}5 \times 10^{-14}$	$2{\cdot}9 \times 10^{-14}$	17×10^{-14}
Plateau basalt	Content	$2{\cdot}2 \times 10^{-6}$	$5{\cdot}0 \times 10^{-6}$	$0{\cdot}80 \times 10^{-2}$	
	Heat generation	$5{\cdot}5 \times 10^{-14}$	$3{\cdot}6 \times 10^{-14}$	$3{\cdot}1 \times 10^{-14}$	12×10^{-14}
Dunite	Content	$1{\cdot}4 \times 10^{-6}$	$1{\cdot}0 \times 10^{-6}$	0	
	Heat generation	$3{\cdot}5 \times 10^{-14}$	$0{\cdot}7 \times 10^{-14}$	0	4×10^{-14}
Eclogite	Content	$1{\cdot}0 \times 10^{-6}$	$0{\cdot}53 \times 10^{-6}$	$0{\cdot}44 \times 10^{-2}$	
	Heat generation	$2{\cdot}5 \times 10^{-14}$	$0{\cdot}4 \times 10^{-14}$	$1{\cdot}7 \times 10^{-14}$	5×10^{-14}

The contents of the elements are in parts by weight; the rates of generation of heat are in calories per second per gram of rock. If we allow for the densities of the rocks, we have the following rates in 10^{-12} calorie per second per cubic centimetre: Granite, 1·3; diorite, etc., 0·83; basalt, etc., 0·50; plateau basalt, 0·36; dunite, 0·13; eclogite, 0·17.

8·22. The importance of radioactivity in rocks was first pointed out by the present Lord Rayleigh in 1906§, following up some less detailed work of Rutherford. The average vertical gradient of temperature in the earth's crust near the surface is about 30° per kilometre. The rate of transfer

* Holmes and Lawson, *Phil. Mag.* **2**, 1926, 1218–1233.
† *Phil. Mag.* **48**, 1924, 819–832; **3**, 1927, 1233–1252.
‡ *Phil. Mag.* **2**, 1926, 1218–1233.
§ *Proc. Roy. Soc.* A, **77**, 1906, 472–485.

of heat across unit surface is the temperature gradient multiplied by the thermal conductivity; the latter, in the sedimentary rocks where the gradients have actually been measured, is probably about 0·008. Thus the rate of transfer of heat outwards to the surface is about $2{\cdot}4 \times 10^{-6}$ cal./cm.^{2} sec., and this would be supplied by the radioactivity of 18 km. of granite, with modern data. (The efficacy of thorium and potassium had not been estimated when Rayleigh wrote.) Of the various possible ways of reconciling the known radioactivity of rocks with the gradient of temperature in the crust, the most probable seemed to be that the radioactivity was confined to a surface layer with a thickness of the order of tens of kilometres. No satisfactory alternative to this hypothesis is known, while several other lines of evidence support it. First, the geological and seismological evidence indicates a passage from granite to dunite or eclogite within a depth of some tens of kilometres; and from the known radioactivities of these rocks we should infer a reduction of the generation of heat to a tenth of its surface value within just such a transition. Second, granite sometimes appears to have been fused several times at some depth, and to have risen towards the earth's surface. In such cases we obtain samples of the upper layers of the magma at later and later stages; and it is found that the radioactive constituents are more abundant in the later specimens, indicating that the upper layers become enriched as time goes on. Thus Holmes quotes* the following contents by weight for Finland granites of decreasing age.

	Ra ($\times 10^{-12}$)	Th ($\times 10^{-5}$)	K ($\times 10^{-2}$)
A	2·36	0·87	2·51
B and C	4·60	2·67	3·61
D	6·21	5·85	5·06

Third, it can be shown easily that if the deep layers were as radioactive as the upper ones, the rocky shell could never have solidified; and the seismological evidence is perfectly clear that it is solid.

8·3. We can now return to the effect of radioactivity in the earth during its early history. The total output of heat from radioactive sources will not have varied greatly during the existence of the earth, and part of the heat conducted out of the earth at present is original heat, a relic of its former heated state. Thus the supply of heat to the surface from radioactive sources can never have been more than about 2×10^{-6} calorie per square centimetre per second. On the other hand the earth receives from the sun about 7×10^{-2} cal./cm.^{2} sec., and even at its present moderate temperature it radiates this away as fast as it receives it. It appears that radioactive heating can never have had any noticeable effect on surface temperatures. Further, so long as the earth remained liquid at the surface, the loss of heat was so rapid that radioactivity cannot be supposed to have

* *Geol. Mag.* **63**, 1926, 317–318.

influenced the course of solidification. But as soon as viscosity became great enough anywhere to stop convection currents radioactivity became a controlling influence. The equation of conduction of heat is

$$\rho c \frac{\partial V}{\partial t} = \frac{\partial}{\partial x}\left(k \frac{\partial V}{\partial x}\right) + \frac{\partial}{\partial y}\left(k \frac{\partial V}{\partial y}\right) + \frac{\partial}{\partial z}\left(k \frac{\partial V}{\partial z}\right) + P \quad \text{.........(1)},$$

where V is the temperature, ρ the density, c the specific heat, k the thermal conductivity, and P the rate of generation of heat per unit volume. If we neglect the curvature of the earth and take the axis of x vertically downwards, we can omit the derivatives with regard to y and z. Further, if the value of dV_0/dx given by 8·11 (2) is maintained to great depths, and the conductivity is uniform, we have

$$\rho c \frac{\partial V}{\partial t} = k \frac{\partial}{\partial x}\left(\frac{\partial V_0}{\partial x}\right) + P \quad \text{.........(2)},$$

and the first term on the right is zero if $\partial V_0/\partial x$ is uniform. Thus the solid is heated up by radioactivity as if no conduction was disturbing the process*. The neglected factors seem unlikely to be important. The inference is that if convection currents were stopped radioactive heating would at once step in and raise the temperature till they started again; and if the temperature were high enough for them to exist the new heat would be carried up as fast as it was generated. It seems in fact that the effect of the heating in this case would be that the interior, instead of cooling till its viscosity was high enough to prevent convection currents entirely, would cool slightly less, till the viscosity was high enough to prevent all convection currents except the very feeble ones needed to carry off the new heat.

If the solidification took place first at intermediate depths, the upper portion would cool in the way just described, but how much of it could actually solidify would depend on whether it could conduct away as much heat as was generated below it. In any case the lower portion was still thoroughly liquid when the intermediate one stiffened, and could convey any new heat upwards as fast as it was produced. Thus the liquid adiabatic gradient would be maintained at great depths until the matter that solidified first had cooled far below its melting point; if indeed this ever took place.

8·31. The seriousness of the last proviso is seen when we consider the flow of heat in the ultimate steady state. The rate of increase of temperature with depth near the surface, as we have seen, is about 30° per kilometre. This gradient is enough to carry off the radioactive heat as fast as it is produced, and also some original heat. When the thermal state is steady the rate of flow of heat out of any sphere must be equal to the rate of generation of heat inside it. When radioactive substances were uniformly distributed through the rocky shell, the rate of generation

* My attention was recently called to this feature by Mr C. H. Waddington.

within, say, 100 km. of the surface was a small fraction of the whole, and the gradient of temperature down to this depth was nearly as at the surface. Hence the temperature 100 km. down was 3000° C., far above the melting point of any conceivable rock. This result is obtained on the hypothesis that the transfer of heat down to that depth is by conduction, and the temperature found is such that the transfer at it must be by convection. The conclusions contradict the hypothesis, and we must suppose that if the internal generation of heat was no greater than now, but uniformly distributed, the rocks below a depth of 50 km. or so could never have solidified. Whatever the mechanism of the original cooling, the heat generated internally was so abundant that it could not be conducted out without temperatures above the melting points of rocks, and permanent fluidity would be maintained except in a thin crust. Below this the viscosity might be high, but not high enough to prevent the new heat from being carried up by convection currents, and the temperature could never fall to the melting point.

8·32. With such an origin as we have supposed, the earth's rocky shell must have been nearly uniform in composition originally. On the other hand the radioactive substances are now strongly concentrated towards the outside. The change is to be attributed to such a process as Holmes has indicated for the granites of Finland. So long as convection went on, the tendency of the radioactive elements to move upwards was effective. As they collected towards the top they left the lower layers impoverished, and the transfer could only stop when the temperature gradient at all depths had become so low as to permit solidification. This implies that the concentration would ultimately be almost complete.

8·33. It has been said that an upward concentration of radioactive elements is unlikely, and that they would tend to collect at great depths on account of their high densities*. But density is far from being the only factor that affects the vertical distribution of the constituents of a fluid, and it is probably unimportant when the relevant constituent is too small in quantity to affect the density of its solvent appreciably. Differences of solubility of radioactive materials in the various layers might be important. Holmes's suggestion† that volatile materials expelled from the lower layers carry the radioactive matter upwards by a sort of process of steam distillation seems, however, to be the best constructive attempt yet made to account for the facts. It has been pointed out by Aston‡ that the inert gases play an extraordinarily small part in the constitution of the earth as a whole; their amounts are of the order of a millionth of those of other

* Schuster, Bronson, A. S. Eve and F. D. Adams (*Nature*, **76**, 1907, 269), have shown that radioactivity is not affected by temperatures up to 2500° C. or pressures up to $2{\cdot}6 \times 10^{10}$ dynes/cm.^{2} The latter corresponds to a depth near 80 km.

† *Geol. Mag.* 1915, 64.

‡ *Nature*, **114**, 1924, 786.

elements, even rare ones. This is explicable if we suppose that the primitive fluid earth was too hot to retain an extensive atmosphere. The inert gases, having no possible resting place except in the atmosphere, would escape from the earth's influence. So would more active gases that could not readily form compounds or solutions able to retain them at magmatic temperatures. It is significant in this connexion that rock analyses show an amount of oxygen almost exactly enough to combine with all the other constituents. The present ocean and atmosphere must then have been originally entirely contained within the primitive magma. We should thus expect the water to have been originally fairly uniformly distributed in the rocky shell, and to have been extruded from it during solidification. But the solidification must have taken place at a temperature far above the critical temperature of water (365° C.), so that the water separated in the form of steam. The steam, rising to the surface, would carry up with it any compounds volatile at magmatic temperatures. Uranium and thorium actually have volatile halogen compounds*.

It has been found experimentally that at high temperatures and pressures water and silicates are miscible in all proportions, and if the present ocean was volatilized it would produce the required conditions for perfect mixing, as Dr J. W. Evans has pointed out†. From this argument alone we must suppose that most of the ocean was once within the earth.

8·34. To sum up, the solidification of the earth probably proceeded as follows. So long as the whole was liquid, cooling was nearly uniform throughout, the temperature gradient being the liquid adiabatic. The next stage would depend on whether the rise of melting point with pressure remained uniform down to the bottom of the rocky shell, or decreased considerably in the first few hundred kilometres. (The melting point in this connexion definitely means the temperature that makes the viscosity high enough to prevent convection currents; but this viscosity is so great in masses of planetary size that the temperature would be indistinguishable from the melting point as usually understood.) In neither case would cooling in the first stage quite reach the melting point anywhere, but it would reach a temperature just sufficiently above it for convection currents to be able to carry off the radioactive heat as fast as it was generated, and the difference would be insignificant. This state would be reached first at the bottom of the rocky shell in the former case, and at an intermediate depth in the latter. Cooling from the surface then made the quasi-solid layer extend up to the surface within a few thousand years. If the first layer to stiffen was an intermediate one, a thick layer of liquid would be trapped below it, and could only cool very slowly, if at all.

When the surface layers became stiff the diminished convection cut down the heat supply from below so much that the temperature sank to

* Cf. *Geol. Mag.* **63**, 1926, 524.

† *Observatory*, 1919, 165–167.

near its present value in a time comparable with that needed for a red-hot coal to become capable of being handled. Cooling by conduction then tended to spread downwards to such a depth that the temperature gradient in the truly solid part was enough to carry off the radioactive heat and the liberated latent heat. This depth was probably of the order of 50 km., and at this stage a permanent steady state would have been reached if the radioactive materials had been able to maintain their uniform distribution. But the separation of the granitic and basaltic layers, and of the oceanic water, began about this time, if not earlier, and carried the radioactive elements upwards until they were so concentrated towards the top that convection below was no longer necessary to carry off the new heat. The time needed for this to happen is quite uncertain. Further cooling from the top would lead to thorough solidification.

8·4. We can now proceed to consider the cooling of a solid crust. Anticipating the result that cooling has by this time become considerable down to a depth of the order of 200 km., we can suppose the depth affected great compared with the thickness of the radioactive layer and small compared with the radius of the earth. The latter condition entitles us to treat the problem as one of flow of heat in one dimension, as for the cooling of a flat plate of infinite depth by radiation from the surface. The equation of heat conduction 8·3 (1) can therefore be taken in the form

$$\rho c \frac{\partial V}{\partial t} = \frac{\partial}{\partial x}\left(k \frac{\partial V}{\partial x}\right) + P \qquad \text{.........(1).}$$

From the former condition we can suppose that all the matter concerned, except the relatively thin surface layers, is non-radioactive. Further, our seismological evidence gives us grounds for assuming that there is no serious change in the physical properties between depths of 30 km. and 1200 km., so that assumptions of uniformity below the radioactive layer are legitimate. Thus k, ρ, and c will be taken constant at great depths. Also the initial temperature will be taken to be $S + mx$, where S is the melting point of deep-seated rocks at ordinary pressures and m is the increase of melting point per unit depth given by 8·11 (2); m also will be assumed constant at present.

If we take simply $$V = mx \qquad \text{.........(2),}$$

equation (1) is satisfied already below the radioactive layers. Near the surface we can take

$$k \frac{\partial V}{\partial x} = k_1 m \qquad \text{.........(3),}$$

where k_1 is the conductivity at great depths; and now

$$V = k_1 m \int^x \frac{dx}{k} + \text{constant} \qquad \text{.........(4)}$$

If we determine the constant so as to keep (2) true below the radioactive layers, we have at all depths a complementary function

$$V = mx + m\int_x^\infty \left(1 - \frac{k_1}{k}\right) dx \qquad \text{.........(5)}$$

$$= V_1 \text{ say} \qquad \text{.........(6).}$$

When allowance is made for the presence of the term P, a solution of (1), independent of the time, is given by

$$k\frac{\partial V}{\partial x} = \int_x^\infty P dx \qquad \text{.........(7),}$$

whence

$$V = \int_0^x \frac{1}{k}\int_x^\infty P dx dx = V_2 \qquad \text{.........(8)}$$

is a solution of (1). $V_1 + V_2$ satisfies (1) completely, but not the initial conditions. If we write

$$V = V_1 + V_2 + V_3 \qquad \text{.........(9),}$$

we have

$$\rho c \frac{\partial V_3}{\partial t} = \frac{\partial}{\partial x}\left(k\frac{\partial V_3}{\partial x}\right) \qquad \text{.........(10).}$$

When $t = 0$,

$$V_3 = S - m\int_x^\infty \left(1 - \frac{k_1}{k}\right) dx - \int_0^x \frac{1}{k}\int_x^\infty P dx dx \qquad \text{.........(11).}$$

The third term on the right of (11) is a constant below the radioactive layer. For $\int_x^\infty P dx$ is zero when x is greater than the thickness of this layer, and the double integral is equal to $\int_0^\infty \frac{1}{k}\int_x^\infty P dx dx = S_0$ say. The second term is zero below the radioactive layer. Thus initially

$$V_3 = S - S_0 + \alpha \qquad \text{.........(12),}$$

where α is a quantity differing from zero only within the radioactive layer.

There is no reason to suspect any systematic change of temperature at the outer surface demanding recognition, so we shall adopt a uniform surface temperature as our zero of temperature. Thus $V = 0$ when $x = 0$ and $t > 0$. V_2 vanishes at $x = 0$, and thus when $x = 0$

$$V_3 = -V_1 = -m\int_0^\infty \left(1 - \frac{k_1}{k}\right) dx = S_1 \qquad \text{.........(13)}$$

say. The conditions satisfied by V_3 are such as to specify a new problem, namely that of the cooling of a mass with no internal generation of heat, and initial temperature $S - S_0 + \alpha$, after the temperature at $x = 0$ has been suddenly reduced to the value in (13) and maintained there. The effects of the term in α are negligible, for the heat it represents would be quickly conducted to the surface and lost. In general, in fact, we can neglect the thermal capacity of the upper layer. Thus in this layer we can take

$$k\frac{\partial V_3}{\partial x} = \text{constant} = k_1\beta \qquad \text{.........(14)}$$

say. Thus when x is a small fraction of 200 km.,

$$V_3 = S_1 + \int_0^x \frac{k_1\beta}{k}\,dx$$
$$= S_1 + \beta x - \beta\int_0^x \left(1 - \frac{k_1}{k}\right) dx \qquad \text{.........(15)}.$$

Thus if $\partial V_3/\partial x$ is equal to β just below the upper layer the corresponding values of V_3 at any time are such that linear extrapolation would make V_3 equal to S_1 at depth

$$x_0 = \int_0^\infty \left(1 - \frac{k_1}{k}\right) dx.$$

We now introduce the thermometric conductivity

$$h^2 = k/c\rho \qquad \text{.........(16)},$$

and denote h in the lower layer by h_1. The solution in the lower layer is

$$V_3 = S - S_0 - (S - S_0 - S_1)\left(1 - \operatorname{Erf}\frac{x - x_0}{2h_1 t^{\frac{1}{2}}}\right) \qquad \text{.........(17)},$$

where Erf is the error function, defined by

$$\operatorname{Erf} u = \frac{2}{\sqrt{\pi}}\int_0^u e^{-v^2}\,dv \qquad \text{.........(18)}.$$

It can be verified by direct differentiation that (17) satisfies the differential equation (10). When u is large $\operatorname{Erf} u$ is nearly 1, and $V_3 = S - S_0$; thus cooling does not affect matter at depths great compared with $2h_1 t^{\frac{1}{2}}$; and also V_3 is equal to $S - S_0$ when t is small, as it should be. Finally we have

$$V_3 = (S - S_0)\operatorname{Erf}\frac{x - x_0}{2h_1 t^{\frac{1}{2}}} + S_1\left(1 - \operatorname{Erf}\frac{x - x_0}{2h_1 t^{\frac{1}{2}}}\right) \qquad \text{.........(19)},$$

so that the gradient of V_3 is such that when $x - x_0$ is small extrapolation to $x = x_0$ would make V_3 equal to S_1. Thus (17) or (19) is the required solution. In the lower layer we therefore have

$$V = V_1 + V_2 + V_3$$
$$= mx + S_0 + V_3$$
$$= mx + S\operatorname{Erf}\frac{x - x_0}{2h_1 t^{\frac{1}{2}}} + (S_0 + S_1)\left(1 - \operatorname{Erf}\frac{x - x_0}{2h_1 t^{\frac{1}{2}}}\right) \qquad \text{.........(20)}.$$

In the upper layers V can be found from (6), (8), and (15). We need chiefly, however, only the gradient of V at the surface. This is

$$\left(\frac{\partial V}{\partial x}\right)_{x=0} = m\frac{k_1}{k} + \frac{1}{k}\int_0^\infty P\,dx + \frac{k_1}{k}(S - S_0 - S_1)\frac{1}{\sqrt{\pi}}\cdot\frac{1}{h_1 t^{\frac{1}{2}}} \qquad \text{.........(21)}.$$

8·41. Numerous exact solutions of problems of the cooling of the earth's crust have been given*, but close approximations to all of them

* Ingersoll and Zobel, *Mathematical Theory of Heat Conduction*, 1913; Holmes, *Geol. Mag.* 1915, 102–112; L. H. Adams, *J. Wash. Acad. Sci.* **14**, 1924, 459–472; Jeffreys, *Phil. Mag.* **32**, 1916, 575–591; *Proc. Roy. Soc.* A, **100**, 1921, 122–149; *Gerlands Beiträge*, **18**, 1927, 1–29; *Operational Methods in Mathematical Physics*, 1927, Chapter V.

can be made on the hypotheses of the last section, and within the limits of accuracy of these approximations the present solution includes all previous ones. The actual datum is provided by equation (21), for $(\partial V/\partial x)_{x=0}$ is the vertical gradient of temperature observed in mines and borings, and (21) is practically an equation to determine $\int P dx$. Unfortunately reliable data concerning the fundamental quantity k are scanty, though they could be obtained easily. The temperature gradient is nearly always measured in the sedimentary rocks at small depths, and therefore the conductivities of these are required. Conductivities of sedimentary rocks are given in some of the standard tables, but these refer to dried specimens, and the conductivities of the rocks *in situ*, containing their natural moisture, are probably considerably greater. If the conductivities of the rocks of the mines and borings where the crustal temperature gradients have been measured were determined in the laboratory, care being taken to preserve the original moisture, the utility of the existing temperature gradients would be considerably enhanced. Dr Ezer Griffiths has pointed out to me privately that it would be even better to measure the conductivities of the rocks while still *in situ*, which would be quite possible. Kelvin* gives 0·0001149 for the conductivity of sandstone *in situ*, equivalent to 0·010 in c.g.s. units. That of granite is about 0·006; we shall take that of a typical sedimentary rock as 0·008. For basalt and dunite we take $k = 0{\cdot}004$. With the latter datum, combined with a specific heat of 0·20 and a density of 3·3, we find $h_1^2 = 0{\cdot}006$. The gradients of temperature at the surface vary from place to place, but a good average value is 0°·00032 C. per centimetre, or 32° per kilometre. With $S = 1400°$ and $t = 1{\cdot}6 \times 10^9$ years $= 5 \times 10^{16}$ sec., the first term on the right of (21) is 2°·0 per kilometre, and the third is rather less than 2° per kilometre. Thus

$$\int_0^\infty P dx = 0{\cdot}008 \times 28{\cdot}0 \times 10^{-5} = 2{\cdot}24 \times 10^{-6}\,\text{cal./cm.}^2\,\text{sec.} \quad (22).$$

If we assume on the basis of the seismological evidence that the intermediate layer is twice as thick as the granitic one, and neglect the radioactivity of the sedimentary and lower layers, which is certainly small, we have from the data in 8·21,

$$\int_0^\infty P dx = 1{\cdot}3 \times 10^{-12} H + 0{\cdot}36 \times 10^{-12} \times 2H = 2 \times 10^{-12} H \quad (23),$$

where H is the thickness of the granitic layer in centimetres. Comparing this with (22) we have

$$H = 11\ \text{km.} \qquad \text{.........(24).}$$

Thus the thermal evidence corresponds to thicknesses of 11 km. and 22 km. for the granitic and intermediate layers respectively. These are in satisfactory agreement with the results derived from the study of near earthquakes and Love waves.

* *Phil. Trans. Roy. Soc. Edin.* 1860, or *Sci. Papers*, 3, 289.

8·411. After Rayleigh's result that the earth's thermal output could be maintained by a thickness of average granite measured in tens of kilometres, the next great advance in principle was by Holmes, who adopted Ingersoll and Zobel's solution for the cooling of a crust containing radioactive matter distributed according to the law

$$P = Ae^{-ax} \qquad \text{.........(25)},$$

where A and a are constants. In this case $\int P dx$ is A/a, where A is the heat output of average granite. Then the equation corresponding to (21) determined a and hence the decline of radioactivity with depth. The present writer, on the supposition of a uniform granitic layer with no radioactivity below, afterwards used the thermal method to estimate the thickness of this layer. On either hypothesis the reduction of radioactivity with depth was such as to indicate that average granite did not exist below about 15 km. L. H. Adams returned to the exponential distribution, allowed for variation of conductivity in the upper layers, and introduced the approximation equivalent to neglecting the heat capacity of the upper layers. It is now possible to bring the theory into close accord with the results of seismology.

8·412. We have when $H < x < 3H$,

$$\int_x^\infty P dx = 0{\cdot}36 \times 10^{-12}(3H - x); \quad \frac{1}{k}\int_x^\infty P dx = 0{\cdot}9 \times 10^{-10}(3H - x) \quad \text{(26)},$$

and when $x < H$,

$$\int_x^\infty P dx = 0{\cdot}72 \times 10^{-12}H + 1{\cdot}3 \times 10^{-12}(H - x) \qquad \text{.........(27)},$$

$$\frac{1}{k}\int_x^\infty P dx = 1{\cdot}2 \times 10^{-10}H + 2{\cdot}2 \times 10^{-10}(H - x) \qquad \text{.........(28)},$$

where k has been taken equal to 0·004 and 0·006 in the intermediate and upper layers respectively. Hence

$$V_2 = \int_0^x \frac{1}{k}\int_x^\infty P dx\, dx = 3{\cdot}4 \times 10^{-10}Hx - 1{\cdot}1 \times 10^{-10}x^2 \quad x < H \qquad \text{......(29)}$$

$$= 2{\cdot}3 \times 10^{-10}H^2 \qquad x = H,$$

and

$$V_2 = 2{\cdot}3 \times 10^{-10}H^2 + 1{\cdot}8 \times 10^{-10}H(x - H) - 0{\cdot}45 \times 10^{-10}(x - H)^2$$

$$H < x < 3H \qquad \text{.........(30)}$$

$$= 4{\cdot}1 \times 10^{-10}H^2 \qquad x = 3H \qquad \text{.........(31)}.$$

With our value of H, V_2 is equal to 280° at the foot of the granitic layer and 490° at that of the intermediate layer. The latter quantity is our S_0.

Again,
$$x_0 = \int_0^\infty \left(1 - \frac{k_1}{k}\right) dx \qquad \text{.........(32)}.$$

With our data the integrand is zero except in the granitic layer, where it is $\frac{1}{3}$. Hence $x_0 = 4$ km. approx. Also

$$S_1 = -mx_0 = -11° \qquad \text{.........(33)},$$

and

$$2h_1 t^{\frac{1}{2}} = 340 \text{ km.}$$

In all, then,

$$V = mx + 480° + 920° \operatorname{Erf} \frac{x - x_0}{340} \qquad \text{.........(34)},$$

where $x - x_0$ is now to be measured in kilometres. m is 3° per kilometre.

8·42. The correctness of this solution depends on the data, and it is desirable to pay some attention to the possibility of important changes in them. The fundamental principle is that the vertical gradient of temperature near the surface is equal to $\frac{1}{k}\int_0^\infty P dx$, apart from the terms in m and $S - S_0$, which are small enough to be treated as corrections. The heat generated in the radioactive layers practically determines the outflow from the surface. Three factors clearly affect the observed gradient: the conductivity of the sedimentary rocks where the gradient is measured; the radioactivity of the local rocks; and the vertical distribution of the radioactivity. All of these vary from place to place, and in the circumstances it is curious that the observed gradient varies so little. Thus Daly* gives the increase of depth corresponding to a rise of temperature of 1° C., for 18 stations in Europe and 14 in North America; those for Europe range from 27·4 to 37·8 metres, with an average of 31·7 metres, and the American ones from 35·7 to 53·7 metres, with an average of 41·8 metres. Apart from the systematic difference between Europe and America, which is hardly surprising, the range in each continent is only in the ratio of about 2 to 3†.

In the paper by Poole and Joly already quoted 13 Finland granites and 9 Hebridean ones are tested. The means for the two regions are as follows:

	Ra	Heat evolution	Th	Heat evolution
Finland	$4{\cdot}7 \times 10^{-12}$	$3{\cdot}6 \times 10^{-13}$	$2{\cdot}9 \times 10^{-5}$	$2{\cdot}1 \times 10^{-13}$
Hebrides	$1{\cdot}8 \times 10^{-12}$	$1{\cdot}4 \times 10^{-13}$	$1{\cdot}1 \times 10^{-5}$	$0{\cdot}8 \times 10^{-13}$

The units are as in the table on p. 143. Allowing in each case an extra $1{\cdot}3 \times 10^{-13}$ calorie per gram per second for potassium, we have the totals $7{\cdot}0 \times 10^{-13}$ for Finland and $3{\cdot}5 \times 10^{-13}$ for the Hebrides. The heat generations of the granites in these two regions differ systematically as 2 : 1,

* *Our Mobile Earth*, 1926, 112–113.

† Temperatures in borings in South Africa have been discussed in relation to the conductivity of the rocks traversed, by H. Pirow, *J. Chem. Metallurg. and Mining Soc. of South Africa*, **25**, 1924, 66–85. The gradient in granite at Dubbeldevlei was 22° per kilometre, distinctly less than European values; but the interpretation of the difference would require more borings and determinations of radioactivity in South Africa and more determinations of conductivity in Europe.

apart from variations within the regions themselves. The estimate given by Holmes and Lawson was a general mean over the earth.

The thickness of the granitic layer probably varies considerably within the continents; beneath mountainous regions it may be twice as thick as beneath plains. The conductivity of the surface rocks varies by an unknown amount. Yet the temperature gradient near the surface, though depending on all these three factors, varies within a continent only as 2 : 3.

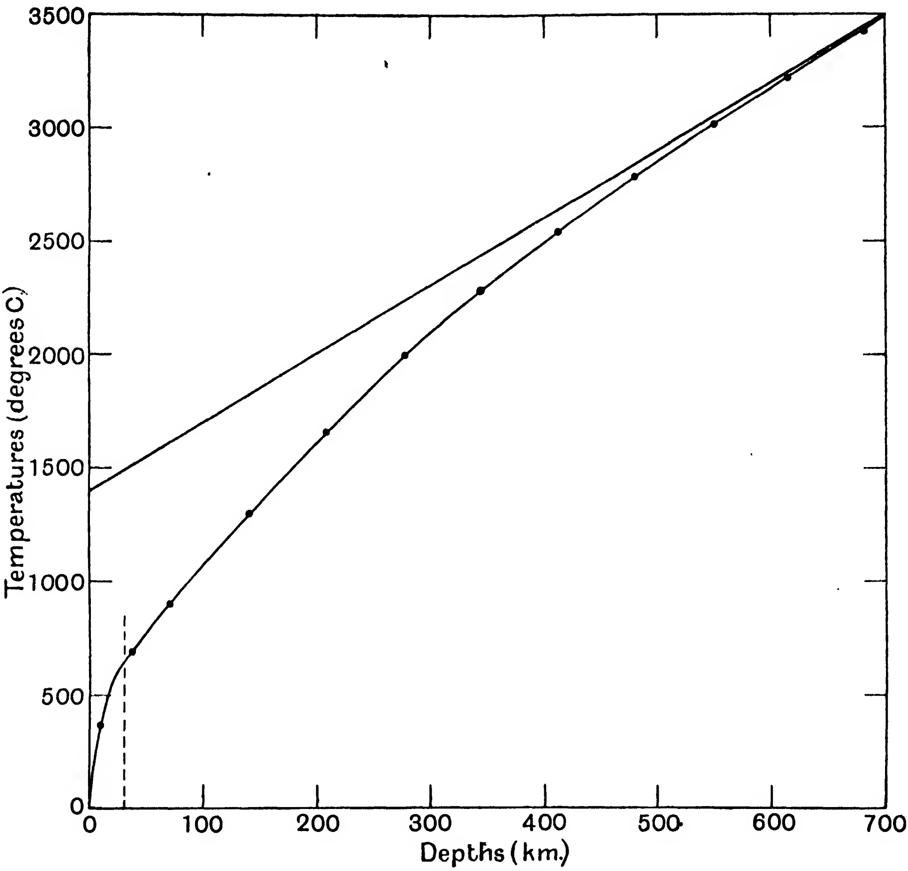


Fig. 13. Distribution of temperature with depth. The straight line represents the function $S + mx$.

It seems to me that this comparative constancy can be attributed only to a correlation between the determining factors. Some of the variation in $(\partial V/\partial x)_{x=0}$, that is, is to be attributed to variation in conductivity, and the rest to variation in $\int P dx$; but the latter integral varies less than either the thicknesses of the upper layers or the radioactivity of the accessible granites separately. The difference between Finnish and Hebridean granites, on this view, would be attributed to greater concentration of radioactivity to the top in Finland, and would not necessarily imply much

difference in the total amount. The radioactivity of the average granite of the upper layer would then be expected to be less than that of the Hebridean granites, whereas the standard value of Holmes and Lawson is rather more (4·9 as against 3·5). On the other hand we cannot suppose average granite less radioactive than even surface diorite, and Holmes and Lawson's value, if excessive, is not likely to be so by as much as 30 per cent. of itself. The value adopted here for the intermediate layer is theirs for plateau basalt; the great outflows of the Deccan and other regions, covering areas of the order of 10^6 square kilometres, are considered by them to be more representative of basalt in depth than the ordinary intrusions, which may have been enriched on the way up. The plateau basalts have lower radioactivities than ordinary ones; if the latter are really typical of the intermediate layer, this layer is more radioactive than we have assumed. On these two grounds the calculations made here may need alteration, but they act in opposite directions, and if they are both adopted the estimated thicknesses of the layers will remain nearly as before.

The determinations of the radioactivity of dunite and eclogite cannot be expected to be directly applicable to the lower layer, for neither of these rocks, as it occurs at the surface, is in its primitive form. With regard to dunite, N. L. Bowen* has given reasons to believe that many rocks near the surface containing much olivine are not formed by the solidification of a lava of the same composition, but by the crystallization of olivine from a lava of basaltic composition. The solid olivine settles to the bottom, where it forms an ultra-basic rock such as dunite or peridotite. The interpretation to be put on the low radioactivity of dunites exposed at the surface is therefore that in the crystallization of such a magna little of the radioactive matter present associates itself with the olivine. So far as it goes it confirms the opinion that the radioactive elements tend towards the top, but does little else.

Eclogite appears to be usually derived from rocks of basaltic composition by metamorphism at high pressures. As was pointed out by Fermor, the fact that its constituents all have high densities would favour its being the form assumed under such conditions. In addition some eclogites are probably formed by the direct crystallization of a basaltic magma at great depths. But in either case its low radioactivity is difficult to understand: the original basalt or basaltic magma should have had the radioactivity of an ordinary basalt, and no way of losing it is provided. The question does not, however, seem to be of importance in our present problems.

Whatever the lower layer may be made of, its radioactivity must be inappreciable; otherwise it could never have solidified, and the temperature gradient in the crust would be greater than is observed. If it is

* *Am. J. Sci.* **14**, 1927, 89–108.

eclogite, the surface basalts and gabbros formed from it must be supposed to have been enriched in radioactive matter in their ascent.

Holmes's hypothesis that the intermediate layer is diorite would imply smaller thicknesses of the layers, perhaps too small to agree with the seismological evidence, unless indeed deep-seated diorite is no more radioactive than ordinary basalt. The opinion of the continental seismologists that the upper layer is about 60 km. thick could be reconciled with the thermal data only if the radioactivity of average granite down to that level was under a quarter of that of average surface granites, that is, under that of plateau basalt.

It appears therefore that the thermal argument concerning the structure of the upper layers is probably trustworthy within, say, about 20 per cent. In comparison with the seismological methods it suffers from the defect of giving only one datum, and therefore cannot by itself determine the thicknesses of the two upper layers separately. The uncertainty of its results in other respects appears to be comparable with that of the methods based on near earthquakes and surface waves; the results are of course consistent. It is capable of much wider application than it has yet received. One would like to know, for instance, whether the lower temperature gradients observed in North America are due to higher conductivity in the sedimentary layer, or lower radioactivity or smaller thicknesses of the upper layers. All these possibilities could be tested without great difficulty, certainly much less than the establishment of the requisite number of reliable seismological stations would require.

8·43. The estimates of the temperatures within the earth are affected by additional uncertainties, because they are largely determined by the quantities m, S, and S_0. In our calculation m has been assumed constant, but actually it may decrease appreciably in the first 500 km. But in any case m is small and its variability can hardly produce an important flow of heat. The *change* of temperature since solidification will not be much affected by variation in m, and it is chiefly this change that will concern us.

S is the temperature of dunite when it acquired a viscosity great enough to stop convection currents, extrapolated to atmospheric pressure. In a pure substance with a sharp melting point S would be equal to the melting point. In a mixed rock, S is at most the melting point of the most fusible mineral; it may be lower than this, because mutual solution at the interfaces between minerals may begin at a lower temperature than would melt any constituent separately. Similarly, when a magma cools, it is not stiff enough for our purpose until the last constituent is solid, and the relevant temperature may be rather low, on account of the presence of volatile constituents, especially water. As successive minerals crystallize, the volatile compounds tend to remain in the magma, and may lower its

melting point considerably. The usual (not universal) order of crystallization appears to be that expressed by H. Rosenbusch*, namely, ores and oxides first, then the ferromagnesian minerals (olivine, augite, etc.), then the felspars, and lastly quartz. The existence of silica in the form of quartz in granite indicates that the rock crystallized below 800° C.; in some rocks (not ordinary granites) the form of the quartz indicates† that it actually crystallized below 575°. The maintenance of quartz in the liquid state far below its normal melting point is to be attributed to the concentration of the volatile constituents in it. But when a granitic magma solidifies near the surface these constituents are expelled, and when it is heated again it does not soften below a temperature of 1150° or so, becoming liquid about 1240°. Basaltic material behaves similarly; Day‡ estimates that the Kilauea lava preserves some mobility down to about 600°, though when it has solidified it does not flow under its own weight till it is reheated to about 1300°. Lower values, near 1000°–1100°, are given for other basalts.

We have to consider how far such phenomena are relevant to our adopted value of S. It must be noticed first of all that the granitic and intermediate layers are together only about 30 km. thick, and that there is no sudden change of properties between depths of 30 km. and 2900 km. The presumption is that this lower layer is approximately uniform in character. Further, whatever its composition may be, there is so much of it that if the two upper layers were mixed with it a refined chemical analysis would be necessary to detect any difference. In fact if the lower layer is now a dunite, it was also a dunite in the earth's earliest stages. S would then be the temperature when separation of olivine crystals from a dunite magma had gone so far as to make a coherent mass. The small amount of material left over when this was no longer an adequate description of the process may have formed the upper layers. The solidification of the lower layer is therefore determined by the crystallization of olivine from a nearly pure magma. Now J. H. L. Vogt§ gives the following temperatures when magmas of the respective types begin to deposit crystals: dunite, 1500°–1600°; gabbro, 1250°; diorite, 1200°; syenite, 1100°; granite, 1000°. The effects of admixture of small quantities of other substances on the melting points of several minerals are known‖. The depression of the melting point per gram molecule of the solute per 100 grams of the mineral is about 400° for orthoclase and 640° for anorthite. The solute being water, of molecular weight 18, the great reductions of melting point found for quartz in granites and other rocks must be taken to imply an amount of

* Quoted by Clarke, *Geochemistry*, 1924, 308.

† Clarke, *loc. cit.* 363.

‡ *J. Frank. Inst.* **200**, 1925, 161–182.

§ *Econ. Geol.* **21**, 1926, 207–233.

‖ Boeke and Eitel, 85. The theory is given in books on thermodynamics, for instance that of Birtwistle, 1927, 115–116.

water of the order of 30 per cent. of the quartz, or, say 6–10 per cent. of the granite as a whole. Since on our hypothesis this is merely the material left over after the concentration of the olivine downwards, the original concentration of water in the rocky shell as a whole can hardly be expected to have exceeded a fraction of 1 per cent. A depression of the melting point by 40° for the latter seems to be the maximum allowable. It seems unlikely, therefore, that our adopted value of 1400° is too high.

The value of S_0, the ultimate steady temperature below the radioactive layers, is somewhat sensitive to variations in the vertical distribution of radioactive matter, even without change in its total amount. Thus if we were to take the whole radioactive layer to be average basalt, keeping the total output of heat the same, the depth required would be 130 km., and S_0 would be 3500°. This is a very extreme instance, but with moderate variations in the distribution S_0 can be changed by about 100°. Also the sedimentary layer, though its low radioactivity may make little contribution to the gradient, may affect S_0; the rise of temperature within this layer must be added to estimates of S_0 because it is included in the second integration. This may give a rise of 100° if the sedimentary layer is 3 km. thick.

The form $S + mx$ for the initial temperature would not hold near the surface, partly through differences in the melting point from several causes, and partly on account of the effects mentioned in 8·3; but this is of minor importance, for the initial temperature within these layers hardly affects the temperature after a long time has elapsed.

8·44. Allusion should be made at this stage to the possibility that in some cases a liquid magma may separate into two or more liquid phases, which afterwards solidify separately. Such liquid immiscibility has been suggested as an explanation of the existence of the granitic and basaltic layers. In work at the Geophysical Laboratory it has been found that, generally speaking, silicates in the liquid state mix perfectly freely, both at atmospheric pressure and under high pressure of water vapour*. Vogt expresses similar views. Recently J. W. Greig has found† that in certain cases magmas containing CaO, MgO, or SrO and much silica, when fused at about 1700°, separate into two liquids. The more acidic of these is nearly pure silica. The other has 3 to 5 molecules of SiO_2 to 2 of the metallic oxide. It appears that these magmas are much too acidic to have any relation to the geophysical problem, as Greig points out. Liquid immiscibility is therefore at present only an outside possibility.

8·5. The foregoing discussion concerns only continental conditions. No mine or boring has yet been made in the ocean bottom, so that the

* N. L. Bowen, *J. Geol. Suppl.* **23**, 1915, 9.
† *Am. J. Sci.* **13**, 1927, 1–44, 133–154.

temperature gradient there is unknown; if such an undertaking could be carried out, most valuable information would be obtained. At present we must do as well as we can without it. Daly remarks*: "Not a cubic inch of granitic rock has ever been found in the hundreds of volcanic islands that dot the region of the central Pacific, one-fourth of the area of the whole globe. So far as they go, these facts suggest that the suboceanic crust is composed of rocks decidedly different from the rocks constituting the continents." This fundamental difference in composition is closely connected with the great problem of the origin of the division of the earth's surface into continents and ocean basins, which has not yet received any convincing explanation. But if we simply accept as a fact the absence of the chief radioactive layer we must expect associated differences in the thermal history. Oceanic basalts do not appear to differ systematically from continental ones in their radioactivity. The basaltic layer being now the top one, and its thickness about 20 km., say, we should expect $\int P dx$ to be half what it is under the continents, and S_0 is found to be about 75°. On the other hand the upward concentration of radioactive matter may have stopped below the oceans when S_0 reached a definite value about equal to its value below the continents; the cooling at great depths would then be nearly the same in continental and oceanic conditions, even though the total output of heat was less; the difference in total output might be counteracted by a distribution through a greater depth. Such a suggestion as this would fit the frequency of oceanic volcanoes. The rise of temperature with depth in the basaltic layer would still be less than in the continents, and lower temperatures are therefore to be expected at equal depths of the order of tens of kilometres.

8·6. It is possible that the actual cooling may involve change of state, and if so the latent heat of fusion will affect the law of cooling. Investigation of its effects has been begun†, but definite results require more knowledge than we have of the variation of latent heat with depth. If it remains constant at about 100 calories per gram, the change of state would be expected to be complete down to a depth of about 300 km., but if it falls this depth will be greater. Seismology indicates no sudden change of properties between 30 km. and 2900 km. down, apart perhaps from the second order discontinuity at 1200 km.

8·7. *Summary.* An account of the method of solidification of the earth has been developed, and has been already summarized in 8·34. The radioactive matter, if originally distributed uniformly, would have a profound influence on the method, and would keep the interior fluid below a skin a few tens of kilometres thick until the processes tending to lift radioactive constituents to the top had time to produce the present concen-

* *Our Mobile Earth*, 98.

† *Gerlands Beiträge*, **18**, 12–15.

tration into the upper layers. A revised theory of the cooling by conduction since solidification has been given. It leads to results concerning the thicknesses of the granitic and intermediate layers consistent with those derived from near earthquakes. The cooling below these layers amounts to about 800° at a depth of 30 km., falling to a few degrees at 600 km. Cooling below the ocean may be the same or more at such depths, but the temperature gradient near the surface below it is probably much less than under the continents, and suboceanic temperatures must be lower than subcontinental ones at depths of tens of kilometres.

CHAPTER IX

The Equations of Motion of an Elastic Solid with Initial Stress

"Lasciate ogni speranza, voi ch' entrate." DANTE, *Inferno*.

9·1. In the following pages several results in the theory of elasticity will be utilized, and it will be convenient to discuss them as far as possible consecutively. If a stress is applied to any substance but a perfect fluid, and we consider a small element of surface within the substance, the stress across the element (i.e. the force per unit area exerted by the matter on one side of it upon the matter on the other side) has three components, one of which is normal to the element, while the other two are tangential to the element. In a perfect fluid the two tangential components are zero.

Consider then a small parallelepiped within the substance, its centre being at the point (x, y, z), and its sides being parallel to the coordinate axes and of lengths dx, dy, dz. Let us denote the components of force per unit area across a plane perpendicular to the axis of x by p_{xx}, p_{xy}, p_{xz}. By considering the opposite faces of the parallelepiped in pairs, we easily see that the resultant force on the element due to the stresses has components

$$\left(\frac{\partial p_{xx}}{\partial x} + \frac{\partial p_{yx}}{\partial y} + \frac{\partial p_{zx}}{\partial z}\right) dx\,dy\,dz \qquad \text{.........(1)},$$

and two symmetrical expressions. If ρ is the density and (X, Y, Z) the bodily force per unit mass, the force on the element arising from this is

$$(\rho X, \rho Y, \rho Z)\, dx\,dy\,dz \qquad \text{.........(2)}.$$

The two sets of forces together produce the acceleration of the element. If u, v, w denote its displacements from its initial position we shall therefore have

$$\rho \frac{d^2 u}{dt^2} = \rho X + \frac{\partial p_{xx}}{\partial x} + \frac{\partial p_{yx}}{\partial y} + \frac{\partial p_{zx}}{\partial z} \qquad \text{.........(3)},$$

with two similar equations. In the differentiation with regard to the time the displacement is supposed to have been expressed as a function of the initial coordinates and the time, and the differentiation is a partial one with regard to the time in this system, so as to give the true acceleration of each element.

Now let us consider how the system of stresses p_{xx}, p_{xy}, p_{xz}, ... p_{zz} can arise. Suppose that in the initial state they had the values indicated by the index zero, p^0_{xx}, p^0_{xy}, ... p^0_{zz}. When an element is displaced the stresses on it in general change, and corresponding changes in its size and form occur. Suppose, for the sake of generality, that at the same time an increase of temperature V takes place, and that the coefficient

of linear expansion is n. Then the alterations in dimensions due to simple expansion in the absence of stress would be such as to make

$$\frac{\partial u}{\partial x}=\frac{\partial v}{\partial y}=\frac{\partial w}{\partial z}=nV;\ \frac{\partial u}{\partial y}=\frac{\partial u}{\partial z}=\frac{\partial v}{\partial x}=\frac{\partial v}{\partial z}=\frac{\partial w}{\partial x}=\frac{\partial w}{\partial y}=0 \quad \text{...(4).}$$

In the actual displacement, however, the values of the displacements will not be such as to satisfy these conditions; the size and form of the element will change on account of the changes in the stresses from neighbouring elements. Now it is proved in works on elasticity that the deformation produced by a small extra stress is related to it according to the laws

$$p'_{xx}=\lambda(e_{xx}+e_{yy}+e_{zz})+2\mu e_{xx},\ \text{etc.} \quad \text{.........(5),}$$

$$p'_{xy}=p'_{yx}=\mu e_{xy}=\mu e_{yx} \quad \text{.........(6),}$$

where accents denote that we are dealing with the changes in the stresses. Here e_{xx}, e_{xy}, e_{xz}, ... e_{zz} are the changes in the quantities

$$\frac{\partial u}{\partial x},\ \frac{\partial v}{\partial x}+\frac{\partial u}{\partial y},\ \frac{\partial w}{\partial x}+\frac{\partial u}{\partial z},\ \ldots,\ \frac{\partial w}{\partial z},$$

measured from the state before the deforming stresses are applied, and λ and μ are the two elastic constants of the element, supposed isotropic. The additional stresses are supposed small enough for their squares to be neglected. Thus we shall have

$$e_{xx}=\frac{\partial u}{\partial x}-nV,\ \text{etc.} \quad \text{.........(7),}$$

$$e_{xy}=\frac{\partial v}{\partial x}+\frac{\partial u}{\partial y},\ \text{etc.} \quad \text{.........(8),}$$

and

$$p'_{xx}=\lambda\left(\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}+\frac{\partial w}{\partial z}\right)+2\mu\frac{\partial u}{\partial x}-(3\lambda+2\mu)nV,\ \text{etc.} \quad \text{......(9),}$$

$$p'_{xy}=\mu\left(\frac{\partial v}{\partial x}+\frac{\partial u}{\partial y}\right) \quad \text{.........(10).}$$

Thus we have for the equations of motion

$$\begin{aligned}\rho\frac{d^2u}{dt^2}=\rho X&+\frac{\partial p^0_{xx}}{\partial x}+\frac{\partial p^0_{yx}}{\partial y}+\frac{\partial p^0_{zx}}{\partial z}\\&+\frac{\partial}{\partial x}\left(\lambda\delta+2\mu\frac{\partial u}{\partial x}\right)+\frac{\partial}{\partial y}\left\{\mu\left(\frac{\partial v}{\partial x}+\frac{\partial u}{\partial y}\right)\right\}+\frac{\partial}{\partial z}\left\{\mu\left(\frac{\partial w}{\partial x}+\frac{\partial u}{\partial z}\right)\right\}\\&-\frac{\partial}{\partial x}\{(3\lambda+2\mu)nV\} \quad \text{.........(11),}\end{aligned}$$

with two symmetrical equations, where δ has been written for

$$\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}+\frac{\partial w}{\partial z} \quad \text{.........(12).}$$

We may also write γ for $\quad (3\lambda+2\mu)nV \quad$ ....(13).

It will be noticed that the bodily force (X, Y, Z) is to be considered as evaluated at (x, y, z) after the deformation. Let (X_0, Y_0, Z_0) be the force at (x, y, z) before the deformation, and put

$$X=X_0+X_1,\ \text{etc.} \quad \text{.........(14).}$$

The stress components $p^0_{xx}, p^0_{xy}, \dots p^0_{zz}$, on the other hand, refer to the particle that is brought to (x, y, z) by the deformation, in its original position and orientation. We need consider only the case where the undisturbed stress is hydrostatic, that is, when complete yield has taken place to any stresses previously existing; other cases of geophysical interest can usually be discussed more easily by approximate methods. If the hydrostatic tension at (x, y, z) before the deformation was p_0, we have for small displacements

$$p^0_{xx} = p_0 - u\frac{\partial p_0}{\partial x} - v\frac{\partial p_0}{\partial y} - w\frac{\partial p_0}{\partial z} \quad \text{.........(15)},$$

$$p^0_{xy} = 0 \quad \text{.........(16)},$$

with corresponding relations for the other components. Also let ρ_0 be the density at (x, y, z) before deformation, and put

$$\rho = \rho_0 + \rho_1 \quad \text{.........(17)}.$$

Then
$$\rho = \rho_0 - \frac{\partial}{\partial x}(\rho_0 u) - \frac{\partial}{\partial y}(\rho_0 v) - \frac{\partial}{\partial z}(\rho_0 w) \quad \text{.........(18)}$$

is the equation of continuity, and determines ρ_1. If the initial state was one of equilibrium, the equations of motion must hold when u, v, w and the additional forces X_1, Y_1, Z_1 are put equal to zero. Thus we have

$$0 = \rho_0 X_0 + \frac{\partial p_0}{\partial x} \quad \text{.........(19)}$$

with two symmetrical relations. Subtracting (19) from (11), omitting terms involving squares and products of the displacements, and using (19) again, we have

$$\rho_0\frac{d^2u}{dt^2} = \rho_0 X_1 + \rho_1 X_0 + \frac{\partial}{\partial x}\left\{\rho_0(uX_0 + vY_0 + wZ_0)\right\}$$
$$+ \frac{\partial}{\partial x}\left(\lambda\delta + 2\mu\frac{\partial u}{\partial x}\right) + \frac{\partial}{\partial y}\left\{\mu\left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}\right)\right\} + \frac{\partial}{\partial z}\left\{\mu\left(\frac{\partial w}{\partial x} + \frac{\partial u}{\partial z}\right)\right\} - \frac{\partial \gamma}{\partial x} \quad \text{...(20)}.$$

In ordinary cases either the displacements (u, v, w) are specified over the boundary, or the displacements and the stress are continuous there. By hypothesis the initial stress was continuous, and the components of the displacement are also continuous, so that the part of the stress arising from the initial stress satisfies the boundary conditions automatically. The remaining portion can be found by the device of considering a small tetrahedron bounded by a small portion of the boundary and by planes parallel to the coordinate planes*; this shows that

$$lp'_{xx} + mp'_{xy} + np'_{xz} \quad \text{.........(21)},$$

with two similar expressions, must be continuous; here (l, m, n) are the direction cosines of the normal to the boundary, and $p'_{xx} \dots$ are to be found from (9) and (10). Since $p'_{xx} \dots$ are small, (l, m, n) can be given their values in the undisturbed state if this is convenient.

* Cf. Love, *The Mathematical Theory of Elasticity*, § 47.

CHAPTER X

The Bending of the Earth's Crust by the Weight of Mountains

"A pupil began to learn geometry with Euclid and asked, when he had learnt one proposition, 'What advantage shall I get by learning these things?' And Euclid called the slave and said, 'Give him sixpence, since he must needs gain by what he learns.'" Sir T. L. HEATH, *A History of Greek Mathematics.*

10·1. In what follows a great deal of attention will be devoted to the mechanism that enables the earth's crust to support the weight of mountains without immediately yielding and gradually effacing all departures from perfect symmetry. A discussion of the stresses in the crust produced by the weight of surface inequalities will therefore be necessary. For this purpose the crust will be supposed to be of finite thickness, and a heavy fluid will be supposed to lie below it. Thus the effect of any additional pressure over the upper surface will be to bend the crust downwards, and the under surface will be forced down into the fluid against hydrostatic pressure. The final deformation will therefore be the result of the joint action of the additional pressures on the upper and lower boundaries.

The problem is a statical one; the accelerations in the equations of motion are zero. There is no change of temperature, so that γ is zero. The elasticity terms in the equations of motion are of order

$$(\lambda + 2\mu)\frac{\partial^2 w}{\partial x^2} \text{ or } (\lambda + 2\mu)\frac{\partial^2 w}{\partial x \partial z},$$

while the gravity terms are comparable with

$$\frac{\partial}{\partial x}(\rho_0 w g) \text{ or } \frac{\partial}{\partial z}(\rho_0 w g).$$

If the wave-length of the disturbing pressure is $2\pi/\kappa$, gravity will be unimportant in comparison with elasticity if $g\rho_0$ is small compared with $(\lambda + 2\mu)\,\kappa$. Now the velocities of the longitudinal waves of earthquakes near the surface enable us to say that $(\lambda + 2\mu)/\rho_0$ is about 5×10^{11} cm.^{2}/sec.^{2} Thus our condition is that κ shall be large compared with 2×10^{-9}/cm., or that $2\pi/\kappa$ shall be small compared with 30,000 km. The effect of bodily gravity will in fact be small in comparison with that of elasticity unless the stresses are of a wave-length comparable with the circumference of the earth, when it would be necessary to include curvature in our calculations as well. Thus the terms depending on bodily forces can be neglected when we are dealing with the support of mountain ranges; the curvature of the earth can be neglected in similar problems; and it will therefore be legitimate to treat the problem as one of simple stress in a uniform flat plate of infinite horizontal extent.

10·2. Let us take the axis of z vertically downwards, and those of x and y horizontal. The origin will be in the middle of the layer, and the thickness of the layer will be $2h$, so that the upper and lower surfaces in their unstrained state are $z = -h$ and $z = h$ respectively. The layer is supposed homogeneous. The equations 9·1 (20) then reduce to

$$(\lambda + \mu)\frac{\partial \delta}{\partial x} + \mu \nabla^2 u = 0 \qquad \text{.........(1)},$$

$$(\lambda + \mu)\frac{\partial \delta}{\partial y} + \mu \nabla^2 v = 0 \qquad \text{.........(2)},$$

$$(\lambda + \mu)\frac{\partial \delta}{\partial z} + \mu \nabla^2 w = 0 \qquad \text{.........(3)}.$$

By differentiating with respect to x, y, and z respectively and adding, we find

$$\nabla^2 \delta = 0 \qquad \text{.........(4)}.$$

Let us write p for $\dfrac{\partial}{\partial x}$, q for $\dfrac{\partial}{\partial y}$, and ϑ for $\dfrac{\partial}{\partial z}$. Put also

$$p^2 + q^2 = -r^2 \qquad \text{.........(5)}.$$

Then
$$\nabla^2 = \vartheta^2 + p^2 + q^2 = \vartheta^2 - r^2 \qquad \text{.........(6)}.$$

Solving for δ by the symbolical method of Heaviside and Bromwich, r being independent of z, we have

$$\delta = A \cosh rz + B \sinh rz \qquad \text{.........(7)},$$

where A and B are functions of x and y only. Let us now introduce a function χ such that

$$\chi = \frac{A}{2r} z \sinh rz + \frac{B}{2r} z \cosh rz \qquad \text{.........(8)}.$$

Then we readily verify by differentiation that

$$(\vartheta^2 - r^2)\,\chi = \delta \qquad \text{.........(9)}.$$

If we substitute for δ in (1), (2) and (3), we have three equations to find u, v, and w from. Particular integrals will evidently be

$$(u_0, v_0, w_0) = -\frac{\lambda + \mu}{\mu}\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right)\chi \qquad \text{.........(10)},$$

since with these values

$$(\lambda + \mu)\frac{\partial \delta}{\partial x} + \mu \nabla^2 u = (\lambda + \mu)\frac{\partial \delta}{\partial x} - \mu(\vartheta^2 - r^2)\frac{\lambda + \mu}{\mu}\frac{\partial \chi}{\partial x} \qquad \text{...(11)},$$

which is zero by (9). The complementary functions are any solutions of

$$(\vartheta^2 - r^2)\,\phi = 0.$$

We have therefore

$$(u, v, w) = -\frac{\lambda + \mu}{\mu}\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right)\chi + (u_1, v_1, w_1) \qquad \text{......(12)},$$

where we can write
$$u_1 = U_0 \cosh rz + U_1 \sinh rz \qquad \text{.........(13)},$$

$$v_1 = V_0 \cosh rz + V_1 \sinh rz \qquad \text{.........(14)},$$

$$w_1 = W_0 \cosh rz + W_1 \sinh rz \qquad \text{.........(15)}.$$

Here U_0, U_1, V_0, V_1, W_0, W_1 are functions of x and y alone. We still have the condition that

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = \delta \qquad \text{.........(16).}$$

Substituting in this from (12), we have

$$pu_1 + qv_1 + \vartheta w_1 = \frac{\lambda + 2\mu}{\mu}\delta \qquad \text{.........(17).}$$

This must hold for all values of z; so equating coefficients of $\cosh rz$ and $\sinh rz$ we have

$$pU_0 + qV_0 + rW_1 = \frac{\lambda + 2\mu}{\mu} A \qquad \text{.........(18),}$$

$$pU_1 + qV_1 + rW_0 = \frac{\lambda + 2\mu}{\mu} B \qquad \text{.........(19).}$$

10·21. We have still to make use of the boundary conditions. These will be supposed to be that there is no shearing stress across the planes $z = \pm h$, and that

when $z = -h$, $\qquad p_{zz} = P + Q \qquad$ ....(20),

when $z = h$, $\qquad p_{zz} = P - Q \qquad$ ....(21).

We have from 9·1 (10)

$$\begin{aligned} p_{zx} &= \mu(pw + \vartheta u) = -2(\lambda + \mu)p\vartheta\chi + \mu(pw_1 + \vartheta u_1) \\ &= -(\lambda + \mu)\frac{p}{r}(A \sinh rz + B \cosh rz) - (\lambda + \mu)pz(A \cosh rz + B \sinh rz) \\ &\quad + \mu\{p(W_0 \cosh rz + W_1 \sinh rz) + r(U_0 \sinh rz + U_1 \cosh rz)\} \text{......(22).} \end{aligned}$$

The vanishing of this where $z = \pm h$ gives the two relations

$$-(\lambda + \mu)\frac{pA}{r}(1 + rh \coth rh) + \mu(pW_1 + rU_0) = 0 \text{......(23),}$$

$$-(\lambda + \mu)\frac{pB}{r}(1 + rh \tanh rh) + \mu(pW_0 + rU_1) = 0 \text{......(24).}$$

The first of these can be written

$$W_1 + \frac{r}{p}U_0 = \frac{\lambda + \mu}{\mu r}(1 + rh \coth rh) A \qquad \text{.........(25),}$$

and we see by symmetry that the vanishing of p_{yz} at the boundaries will prove the right side of (25) equal to $W_1 + \frac{r}{q}V_0$. Similarly (24) gives

$$\begin{aligned} W_0 + \frac{r}{p}U_1 &= \frac{\lambda + \mu}{\mu r}(1 + rh \tanh rh) B \qquad \text{.........(26)} \\ &= W_0 + \frac{r}{q}V_1. \end{aligned}$$

Using these relations to eliminate U_0, V_0, U_1, V_1 from (18) and (19), we find

$$W_1 = \frac{\lambda + 2\mu}{2\mu r} A + \frac{\lambda + \mu}{2\mu r}(1 + rh \coth rh) A \qquad \text{.........(27),}$$

$$W_0 = \frac{\lambda + 2\mu}{2\mu r} B + \frac{\lambda + \mu}{2\mu r}(1 + rh \tanh rh) B \qquad \text{.........(28).}$$

We have also $$p_{zz} = \lambda\delta + 2\mu\vartheta w$$
$$= \lambda\delta - 2(\lambda+\mu)\vartheta^2\chi + 2\mu\vartheta w_1 \qquad \text{.........(29).}$$
But $$\vartheta^2\chi = r^2\chi + \delta.$$
Hence $$p_{zz} = -(\lambda+2\mu)\delta - 2(\lambda+\mu)r^2\chi + 2\mu\vartheta w$$
$$= -(\lambda+2\mu)(A\cosh rz + B\sinh rz)$$
$$-(\lambda+\mu)rz(A\sinh rz + B\cosh rz)$$
$$+2\mu r(W_0\sinh rz + W_1\cosh rz) \qquad \text{.........(30).}$$
By hypothesis p_{zz} is to be equal to $P+Q$ when $z=-h$, and to $P-Q$ when $z=h$. Hence
$$-\{(\lambda+2\mu)+(\lambda+\mu)rh\tanh rh\}A + 2\mu rW_1 = P\operatorname{sech} rh \qquad \text{...(31),}$$
$$-\{(\lambda+2\mu)+(\lambda+\mu)rh\coth rh\}B + 2\mu rW_0 = -Q\operatorname{cosech} rh \ \text{...(32).}$$
Substituting for W_0 and W_1 from (27) and (28), we have
$$(\lambda+\mu)\{1 + rh\operatorname{cosech} rh\operatorname{sech} rh\}A = P\operatorname{sech} rh \qquad \text{......(33),}$$
$$(\lambda+\mu)\{1 - rh\operatorname{cosech} rh\operatorname{sech} rh\}B = -Q\operatorname{cosech} rh \qquad \text{...(34).}$$

These equations, with the expressions for the other coefficients in terms of A and B, give the symbolic solution of the problem when P and Q are known. In particular, if P and Q are proportional to $\sin\kappa x\sin\kappa' y$, where κ and κ' are constants, or to any other product of harmonic functions of κx and $\kappa' y$, the solution is to be found by simply substituting $(\kappa^2+\kappa'^2)^{\frac{1}{2}}$ for r.

10·22. We require to know the stress-differences in the solid layer. The stresses are given by
$$p_{xx} = \lambda\delta + 2\mu\frac{\partial u}{\partial x}, \text{ with two similar expressions;}$$
$$p_{xy} = \mu\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right), \text{ with two similar expressions.}$$
Substituting for the coefficients in these, we find
$$p_{zz} = (\lambda+\mu)\{(1 + rh\coth rh)A\cosh rz + (1 + rh\tanh rh)B\sinh rz$$
$$- rz(A\sinh rz + B\cosh rz)\} \text{.........(35),}$$
$$p_{xx} = A\cosh rz\left(\lambda - \mu\frac{p^2}{r^2} + (\lambda+\mu)\frac{p^2}{r}h\coth rh\right) + B\sinh rz\left(\lambda - \mu\frac{p^2}{r^2}\right.$$
$$\left. + \frac{(\lambda+\mu)p^2}{r}h\tanh rh\right) - \frac{(\lambda+\mu)p^2}{r}z(A\sinh rz + B\cosh rz) \text{ ...(36),}$$
$$p_{zx} = (\lambda+\mu)\frac{p}{r}\{Arh\coth rh\sinh rz + Brh\tanh rh\cosh rz$$
$$- rz(A\cosh rz + B\sinh rz)\} \qquad \text{.........(37),}$$
$$p_{xy} = -(\lambda+\mu)pq\frac{z}{r}(A\sinh rz + B\cosh rz)$$
$$+\frac{pq}{r^2}[\{(\lambda+\mu)rh\coth rh - \mu\}A\cosh rz + \{(\lambda+\mu)rh\tanh rh - \mu\}B\sinh rz]$$
$$\text{.........(38),}$$
with symmetrical expressions for p_{yy} and p_{yz}.

It is easy to verify that these expressions satisfy the original equations of equilibrium

$$\frac{\partial p_{xx}}{\partial x} + \frac{\partial p_{xy}}{\partial y} + \frac{\partial p_{xz}}{\partial z} = 0 \qquad \text{.........(39),}$$

with two similar equations.

10·3. In the problem of the support of the earth's crust, the values of P and Q must satisfy two conditions. Suppose the pressure to be reduced by S over the surface $z = -h$. This surface will be depressed by a certain amount, and the reduction in the weight of the matter above this level must be allowed for in finding the stress after deformation. If then ρ be the density of the superficial matter, we must have

$$P + Q = S + g\rho\,(w)_{z=-h} \qquad \text{.........(1).}$$

Similarly, the depression of the lower boundary by the amount $w_{z=h}$ exposes it to the additional pressure due to a height $w_{z=h}$ of the underlying fluid, whose density will be supposed to be ρ'. Hence the additional stress p_{zz} across the boundary will be increased by $-g\rho' w_{z=h}$. The negative sign is required because a positive stress is a tension, not a pressure. But this is the change in the stress on the matter constituting the boundary; after the displacement the vertical tension across the plane $z = h$ has to support the weight of the column of the crust below it, and therefore the increase of tension there is $g\rho w_{z=h}$ more than that on the matter at the boundary. Hence

$$P - Q = -g\,(\rho' - \rho)\,w_{z=h} \qquad \text{.........(2).}$$

Now we have

$$\begin{aligned} w &= -\frac{\lambda+\mu}{\mu}\frac{\partial\chi}{\partial z} + w_1 \\ &= -\frac{\lambda+\mu}{2\mu} z\,(A\cosh rz + B\sinh rz) \\ &\quad + \frac{\lambda+2\mu}{2\mu r}(A\sinh rz + B\cosh rz) \\ &\quad + \frac{\lambda+\mu}{2\mu}\,.\,h\,(A\coth rh\sinh rz + B\tanh rh\cosh rz) \qquad \text{.........(3).} \end{aligned}$$

Hence

$$w_{z=-h} = \frac{\lambda+2\mu}{2\mu r}(-A\sinh rh + B\cosh rh) \qquad \text{.........(4),}$$

$$w_{z=h} = \frac{\lambda+2\mu}{2\mu r}(A\sinh rh + B\cosh rh) \qquad \text{.........(5).}$$

On substituting these expressions in (1) and (2), and expressing P and Q in terms of A and B by means of 10·21 (33) and (34), we have

$$(\lambda+\mu)(A\cosh rh + Arh\,\text{cosech}\,rh - B\sinh rh + Brh\,\text{sech}\,rh)$$
$$= S + g\rho\frac{\lambda+2\mu}{2\mu r}(-A\sinh rh + B\cosh rh) \qquad \text{.........(6),}$$

$$(\lambda+\mu)(A\cosh rh + Arh\,\text{cosech}\,rh + B\sinh rh - Brh\,\text{sech}\,rh)$$
$$= -g\,(\rho'-\rho)\frac{\lambda+2\mu}{2\mu r}(A\sinh rh + B\cosh rh) \qquad \text{.........(7),}$$

which suffice to determine A and B. If we write

$$A - B = 2M;\ A + B = 2N \qquad \text{.........(8)},$$

(6) and (7) take the alternative forms

$$(\lambda+\mu)\left\{Me^{rh} + Ne^{-rh} + \frac{rh}{\cosh rh \sinh rh}(Ne^{rh} + Me^{-rh})\right\}$$
$$= S + \frac{g\rho(\lambda+2\mu)}{2\mu r}(-Me^{rh} + Ne^{-rh}) \qquad \text{.........(9)},$$

$$(\lambda+\mu)\left\{Ne^{rh} + Me^{-rh} + \frac{rh}{\cosh rh \sinh rh}(Me^{rh} + Ne^{-rh})\right\}$$
$$= -g(\rho'-\rho)\frac{\lambda+2\mu}{2\mu r}(Ne^{rh} - Me^{-rh}) \qquad \text{.........(10)}.$$

Let us first consider the case where rh is large. Then (10) shows that Ne^{rh} and $4rhMe^{-rh}$ are of the same order; for we already know that $(g\rho)/(\lambda+2\mu)\,\kappa$ is small in the problems we are considering, and *a fortiori* $\dfrac{g(\rho'-\rho)(\lambda+2\mu)}{2(\lambda+\mu)\mu r}$ is small. Hence to a first approximation

$$N = -4rhMe^{-2rh} \qquad \text{.........(11)},$$

and substituting in (9) we have

$$M = Se^{-rh}/(\lambda+\mu) \qquad \text{.........(12)}.$$

Hence by 10·2 (7)
$$\delta = \frac{S}{\lambda+\mu}\{e^{-r(z+h)} - 4rhe^{-3rh+rz}\} \qquad \text{.........(13)},$$

or if we take as a new coordinate the depth below the undisturbed surface, putting

$$z + h = z_1 \qquad \text{.........(14)},$$

and neglect the terms in e^{rz}, on account of their small coefficients, we find from 10·22 (35) to (38)

$$p_{zz} = (\lambda+\mu)\,rA\,(h+z)\,e^{-rz} + (\lambda+\mu)\,Ae^{-rz} \qquad \text{.........(15)},$$

$$p_{xx} = \left(\lambda - \frac{\mu p^2}{r^2}\right)Ae^{-rz} + \frac{(\lambda+\mu)\,p^2}{r}A\,(h+z)\,e^{-rz} \qquad \text{...(16)},$$

$$p_{zx} = -(\lambda+\mu)\,pA\,(h+z)\,e^{-rz} \qquad \text{.........(17)},$$

$$p_{xy} = (\lambda+\mu)\frac{pq}{r}A\,(h+z)\,e^{-rz} - \mu\frac{pq}{r^2}Ae^{-rz} \qquad \text{.........(18)}.$$

These can be shown by differentiation to satisfy the conditions of equilibrium. We have from (8) and (12)

$$(\lambda+\mu)\,Ae^{rh} = S \qquad \text{.........(19)}.$$

Hence the stresses reduce to

$$p_{zz} = S\,(1 + rz_1)\,e^{-rz_1} \qquad \text{.........(20)},$$

$$p_{xx} = S\frac{\lambda - \mu p^2/r^2}{\lambda+\mu}e^{-rz_1} + \frac{p^2}{r}z_1e^{-rz_1} \qquad \text{.........(21)},$$

$$p_{zx} = -Spz_1e^{-rz_1} \qquad \text{.........(22)},$$

$$p_{xy} = S\frac{pq}{r^2}\left(rz_1 - \frac{\mu}{\lambda+\mu}\right)e^{-rz_1} \qquad \text{.........(23)}.$$

10·4. If now S is a simple harmonic function of κx and $\kappa' y$, we may fix the origin so that

$$S = \nu \cos \kappa x \cos \kappa' y \qquad \text{.........(1)},$$

where ν is a constant. Then $\quad r^2 = \kappa^2 + \kappa'^2 \qquad$ ....(2),

and

$$p_{zz} = \nu\,(1 + r z_1)\, e^{-r z_1} \cos \kappa x \cos \kappa' y \qquad \text{.........(3)},$$

$$p_{xx} = \nu \left\{1 - \frac{\mu \kappa'^2}{(\lambda + \mu)\, r^2} - \frac{\kappa^2 z_1}{r}\right\} e^{-r z_1} \cos \kappa x \cos \kappa' y \qquad \text{......(4)},$$

$$p_{zx} = \nu \kappa z_1 e^{-r z_1} \sin \kappa x \cos \kappa' y \qquad \text{.........(5)},$$

$$p_{xy} = \frac{\nu \kappa \kappa'}{r^2}\left(r z_1 - \frac{\mu}{\lambda + \mu}\right) e^{-r z_1} \sin \kappa x \sin \kappa' y \qquad \text{.........(6)},$$

from which the expressions for p_{yy} and p_{yz} can be obtained by considerations of symmetry. Now the three principal stresses at any point are the roots of the cubic equation in ϖ,

$$\begin{vmatrix} p_{xx} - \varpi & p_{xy} & p_{xz} \\ p_{xy} & p_{yy} - \varpi & p_{yz} \\ p_{xz} & p_{yz} & p_{zz} - \varpi \end{vmatrix} = 0 \qquad \text{.........(7)}.$$

This determinant is an even function both of $\cos \kappa x$, $\sin \kappa x$, $\cos \kappa' y$, and $\sin \kappa' y$. The stationary values of each stress-component with respect to variations in x and y occur where $\cos \kappa x$ and $\cos \kappa' y$ are each equal to zero or $\pm$ 1. The same may be expected to be true of the principal stresses and of the differences between them, and this has been found to hold in all special cases so far examined; but a general proof has not yet been constructed, on account of the heaviness of the algebra involved. It will, however, be assumed that it is in general true that the greatest stress-differences occur at points where $\cos \kappa x$ and $\cos \kappa' y$ are each zero or $\pm$ 1, and the following discussion will therefore deal only with these points.

It will facilitate discussion if the surface of the earth is considered marked out into rectangles by the two sets of lines where

$$\cos \kappa x = 0 \text{ and } \cos \kappa' y = 0.$$

Then the points of the surface where $\cos \kappa x$ and $\cos \kappa' y$ are $\pm$ 1 will be the centres of these rectangles. If $\cos \kappa x$ and $\cos \kappa' y$ are both 1 or both $-$ 1, the point considered will be one of maximum elevation; if only one of them is $-$ 1, the point will be one of maximum depression. The points where

$$\cos \kappa x = \sin \kappa' y = 0,$$

and those where

$$\sin \kappa x = \cos \kappa' y = 0$$

are the midpoints of the sides of the rectangles. Those where

$$\cos \kappa x = \cos \kappa' y = 0$$

are the corners of the rectangles. We therefore proceed to discuss the stress-differences at points vertically below these various points identified on the surface.

10·41. Considering first the centres of the rectangles, we have

$$p_{yz} = p_{zx} = p_{xy} = 0 \qquad \text{.........(1).}$$

Hence the three principal stresses at any depth are p_{xx}, p_{yy}, p_{zz}. We see at once that the greatest of these in absolute value is always p_{zz}, and the greatest stress-difference is therefore either

$$| p_{zz} - p_{xx} | = \left(\frac{\mu\kappa'^2}{(\lambda + \mu) r^2} + \frac{r^2 + \kappa^2}{r} z_1 \right) \nu e^{-rz_1} \qquad \text{.........(2),}$$

or

$$| p_{zz} - p_{yy} | = \left(\frac{\mu\kappa^2}{(\lambda + \mu) r^2} + \frac{r^2 + \kappa'^2}{r} z_1 \right) \nu e^{-rz_1} \qquad \text{.........(3).}$$

The maximum value of $p_{zz} - p_{xx}$ in a vertical line is given by

$$\frac{\partial}{\partial z_1} \left\{ \left(\frac{\mu\kappa'^2}{(\lambda + \mu) r^2} + \frac{r^2 + \kappa^2}{r} z_1 \right) e^{-rz_1} \right\} = 0 \qquad \text{.........(4),}$$

which makes

$$z_1 = \frac{2\kappa^2 + \kappa'^2\lambda/(\lambda + \mu)}{r(2\kappa^2 + \kappa'^2)} \qquad \text{.........(5).}$$

Similarly the greatest value of $p_{zz} - p_{yy}$ occurs at a depth

$$\frac{2\kappa'^2 + \kappa^2\lambda/(\lambda + \mu)}{r(2\kappa'^2 + \kappa^2)} \qquad \text{.........(6).}$$

So far the magnitudes of κ and κ' have not been considered. Without loss of generality, therefore, we may suppose the axes turned so that the axis of y is in the direction of the greater wave length. Thus κ will be greater than κ'. Then

$$| p_{zz} - p_{xx} | - | p_{zz} - p_{yy} | = (\kappa^2 - \kappa'^2) \left(- \frac{\mu}{(\lambda + \mu) r^2} + \frac{z_1}{r} \right) \nu e^{-rz_1} \ \text{...(7).}$$

Thus $| p_{zz} - p_{xx} |$ will be greater than $| p_{zz} - p_{yy} |$ if the depth exceeds $\frac{\mu}{(\lambda + \mu) r}$, but the latter will be the greater if the depth is less than this critical value. Now the greatest value of $| p_{zz} - p_{yy} |$ occurs at a depth $\frac{2\kappa'^2 + \kappa^2\lambda/(\lambda + \mu)}{r(2\kappa'^2 + \kappa^2)}$. The condition for this to be greater than $\frac{\mu}{(\lambda + \mu) r}$ is

$$(\lambda + \mu) 2\kappa'^2 + \kappa^2\lambda > \mu (2\kappa'^2 + \kappa^2) \qquad \text{.........(8),}$$

which is true, since λ is positive and about equal to μ. Hence the greatest value of $| p_{zz} - p_{yy} |$ occurs at a depth where it is less than the value of $| p_{zz} - p_{xx} |$ at the same depth, and *a fortiori* is less than the value of $| p_{zz} - p_{xx} |$ at the depth where the latter reaches its maximum. Hence the greatest stress-difference below the centre of a rectangle is $| p_{zz} - p_{xx} |$ at the depth $Z = \frac{2\kappa^2 + \kappa'^2\lambda/(\lambda + \mu)}{r(2\kappa^2 + \kappa'^2)}$, its value there being $\frac{2\kappa^2 + \kappa'^2}{r^2} \nu e^{-rZ}$.

10·42. At the corners of the rectangles we have

$$p_{zz} = p_{xx} = p_{yy} = p_{zx} = p_{yz} = 0 \qquad \text{.........(1),}$$

$$p_{xy} = \pm \frac{\nu\kappa\kappa'}{r^2} \left(rz_1 - \frac{\mu}{\lambda + \mu} \right) e^{-rz_1} \qquad \text{.........(2).}$$

Hence $$\varpi = 0 \text{ or } \pm \frac{\nu\kappa\kappa'}{r^2}\left(rz_1 - \frac{\mu}{\lambda+\mu}\right)e^{-rz_1} \qquad \text{.........(3)},$$

and the greatest stress-difference is

$$\frac{2\nu\kappa\kappa'}{r^2}\left|rz_1 - \frac{\mu}{\lambda+\mu}\right|e^{-rz_1} \qquad \text{.........(4)}.$$

10·43. To compare this with the greatest stress-difference at an equal depth below a point of maximum elevation or depression, we must consider two separate cases, according as z_1 is greater or less than $\frac{\mu}{(\lambda+\mu)r}$. If

$$z_1 > \frac{\mu}{(\lambda+\mu)r},$$

the stress-difference at a corner is less than $\frac{2\nu\kappa\kappa'}{r}z_1e^{-rz_1}$. The coefficient of $\nu z_1 e^{-rz_1}$ in this is $\frac{2\kappa\kappa'}{r}$, which is less than $\frac{2\kappa^2}{r}$, which is less than $\frac{2\kappa^2+\kappa'^2}{r}$, which is the coefficient of $\nu z_1 e^{-rz_1}$ in the expression for $|p_{zz} - p_{xx}|$ vertically below the centre of a rectangle. The other term in the latter expression is always positive, and therefore the stress-difference at a corner at any depth greater than $\frac{\mu}{(\lambda+\mu)r}$ is less than the stress-difference at an equal depth below a centre.

If the depth is less than $\frac{\mu}{(\lambda+\mu)r}$, the stress-difference below a corner is not greater than $\frac{2\nu\kappa\kappa'\mu}{r^2(\lambda+\mu)}$, and attains this value at the surface. Below a centre $|p_{zz} - p_{yy}|$ is greater than $|p_{zz} - p_{xx}|$; it is at least equal to $\frac{\mu\kappa^2}{r^2(\lambda+\mu)}$, attaining this value at the surface. Hence if κ is greater than $2\kappa'$, the stress-difference at the corner at any depth is less than that at the centre at the same depth; but if κ is less than $2\kappa'$, the stress-difference at the corner is the greatest at the surface.

Let us now compare the stress-difference at the surface at a corner, when κ is less than $2\kappa'$, with the greatest stress-difference at any depth below a centre. The maximum stress-difference below a centre is

$$\frac{2\kappa^2+\kappa'^2}{r^2}\nu e^{-rZ}.$$

But the expression for Z shows that rZ is less than unity, and therefore this stress-difference is greater than $\frac{2\kappa^2+\kappa'^2}{r^2 e}\nu$. The condition for this to be greater than that at the surface at a corner is therefore that

$$2\kappa^2+\kappa'^2 > 2e\kappa\kappa'\frac{\mu}{\lambda+\mu}.$$

If λ and μ are equal, this reduces to

$$2\kappa^2 - e\kappa\kappa' + \kappa'^2 > 0,$$

which is true, since $e^2 < 8$. Hence if the Poisson ratio of the matter of the earth's crust has the standard value of $\frac{1}{4}$, which makes λ equal to μ, the greatest stress-difference below a centre will be greater than the greatest at a corner.

10·44. Considering now the midpoints of the sides, we have if

$$\cos \kappa x = \sin \kappa' y = 0,$$

$$p_{zz} = p_{xx} = p_{yy} = p_{yz} = p_{xy} = 0 \qquad \text{.........(1)},$$

$$p_{zx} = \pm\, \nu\kappa z_1 e^{-rz_1} \qquad \text{.........(2)}.$$

Hence

$$\varpi = 0 \text{ or } \pm\, \nu\kappa z_1 e^{-rz_1} \qquad \text{.........(3)},$$

and the greatest stress-difference is $2\nu\kappa z_1 e^{-rz_1}$. A sufficient condition for this to be less than the value of $|\, p_{zz} - p_{xx} \,|$ at the same depth below a centre is that 2κ shall be less than $(r^2 + \kappa^2)/r$. On substituting for r in terms of κ and κ' we find that this becomes

$$4\kappa^2 (\kappa^2 + \kappa'^2) < (2\kappa^2 + \kappa'^2)^2 \qquad \text{.........(4)}$$

which is true unless κ' is zero. Thus the stress-difference below the midpoint of a side of this set is always less than that at the same depth below the centre. Since κ' is less than κ, we see that the stress-difference at the midpoint of a side of the other set is still smaller.

10·45. Collecting these results, we see that if κ is greater than $2\kappa'$, so that the wave-length in one direction is more than twice that in the other direction, the stress-difference at any point vertically below a place of maximum elevation or depression is greater than that at the same depth anywhere else. The maximum value is at a depth Z, equal to $\dfrac{2\kappa^2 + \kappa'^2\lambda/(\lambda + \mu)}{r\,(2\kappa^2 + \kappa'^2)}$, and its amount there is $\dfrac{2\kappa^2 + \kappa'^2}{r^2}\,\nu e^{-rZ}$, and greater than $\dfrac{2\kappa^2 + \kappa'^2}{r^2 e}\,\nu$. If, however, κ is between κ' and $2\kappa'$, there will be regions of finite depth, around the points where the lines of zero elevation cross, where the stress-difference exceeds that at the same depth at the points of maximum elevation or depression. Even in this case, however, it will not with normal substances exceed the greatest stress-difference vertically below a point of maximum elevation. Thus if the strength of the material is the same at all depths, permanent deformation will take place first by a sinking of the greatest elevations and a rise of the greatest depressions. Since the greatest stress-difference in these circumstances is $|\, p_{zz} - p_{xx} \,|$, the failure will take place by the greater stress p_{zz} overcoming the smaller p_{xx}, so that the matter below a region of maximum elevation will be forced out parallel to the axis of x; in other words, in the direction of the shorter wave-length. Flow in the direction of the longer wave-length will occur only if the surface stress is applied so quickly that the stress-difference $|\, p_{zz} - p_{yy} \,|$ reaches the amount necessary to produce

deformation before flow in the direction of the shorter wave-length has had time to reduce the stress-difference $|p_{zz} - p_{xx}|$ below the limit necessary to produce yield. Flow in regions below the slopes will become possible only when extra stress has been thrown on these regions by the failure of materials below the summits and hollows to support their share.

If κ is between κ' and $2\kappa'$, and the matter near the surface is weaker than that below, the stress-difference at the corners near the surface may become enough to produce yield before that below the centres has reached the greater value necessary to produce yield at the greater depth. The same may happen even in the case of uniform strength if μ is greater than λ; but in the actual earth these quantities are nearly equal. Now the stress-difference at the corners arises from the component p_{xy}, which represents a tangential shear across each of the planes $x =$ constant, $y =$ constant. The tendency will therefore be for matter near the places of greatest slope to flow downhill.

10·5. The present problem is a generalization of one discussed by Sir G. H. Darwin*, as the limiting case of the stresses due to a spherical harmonic deformation of high order. His results refer only to the case where κ' is zero, corresponding to a series of parallel mountain chains of infinite length. In this case our results 10·4 (3) to (6) reduce to

$$p_{zz} = \nu(1 + \kappa z_1) e^{-\kappa z_1} \cos \kappa x \qquad \text{.........(1)},$$

$$p_{xx} = \nu(1 - \kappa z_1) e^{-\kappa z_1} \cos \kappa x \qquad \text{.........(2)},$$

$$p_{yy} = \nu \frac{\lambda}{\lambda + \mu} e^{-\kappa z_1} \cos \kappa x \qquad \text{.........(3)},$$

$$p_{zx}' = \nu \kappa z_1 e^{-\kappa z_1} \sin \kappa x \qquad \text{.........(4)},$$

$$p_{zy} = p_{xy} = 0 \qquad \text{.........(5)}.$$

The equation giving the principal stresses therefore becomes, if

$$\varpi = \sigma \nu e^{-\kappa z_1} \qquad \text{.........(6)},$$

$$\begin{vmatrix} (1 - \kappa z_1)\cos \kappa x - \sigma & 0 & \kappa z_1 \sin \kappa x \\ 0 & \frac{\lambda}{\lambda + \mu}\cos \kappa x - \sigma & 0 \\ \kappa z_1 \sin \kappa x & 0 & (1 + \kappa z_1)\cos \kappa x - \sigma \end{vmatrix} = 0 \quad \text{...(7)},$$

the roots of which are $\frac{\lambda}{\lambda + \mu}\cos \kappa x$ or $\cos \kappa x \pm \kappa z_1$. The first of these roots evidently corresponds to the component p_{yy}, flow due to which is impossible, since it is the only component capable of producing motion parallel to the axis of y, and is itself independent of y, so that the resultant force parallel to the axis of y acting on any portion of the crust must be zero. The two principal stresses in a vertical plane perpendicular to the ridges differ by $2\nu\kappa z_1 e^{-\kappa z_1}$. This is independent of x, so that the stress-difference is the same at all points at the same depth, whatever their horizontal distances from the nearest ridge. This has been seen to be untrue in the more general case. The maximum stress-difference occurs

* *Scientific Papers*, **2**, 481–484.

at a depth $1/\kappa$, and is equal to $2\nu/e$. There is no stress-difference at the surface, in which respect again we have a disagreement with the case where κ' is not zero.

10·6. In the above discussion rh has been supposed large. The stress-components introduced by the pressure applied over the surface have been found to decrease rapidly with depth, when the depth exceeds a definite finite value of order $1/r$. Hence the fluid upon which the crust has been supposed to rest has practically no additional stress to support, and it was therefore to be expected that the reaction $P - Q$ across the boundary between the solid and the fluid would be negligible, as has been found to be the case. So long as the condition that rh is great is satisfied, the solution for the regions where the stress is comparable with the pressure applied over the surface is the same as that for a solid of infinite depth.

If, however, rh is small, the stresses at all points within the solid layer would be expected to be comparable, and the fluid may have to support a large fraction of the pressure applied over the surface. In other words, the crust may undergo considerable flexure, and the extra hydrostatic pressure where it is depressed into the fluid may nearly balance the pressure on the top. Within the fluid there is no stress-difference, but in the surface layer there will be some. In comparing the two cases it will be useful to evaluate the stress-differences in the crust, so as to find out whether they exceed those in a deep crust or not. Coming now to the case of a floating crust, of small depth in comparison with the wave-length of the deformation applied, we may approximate to 10·3 (9) and (10) by neglecting powers of rh higher than the first. They reduce to

$$2(\lambda + \mu) A = S + g\rho \frac{\lambda + 2\mu}{2\mu r}(-Arh + B) \qquad \text{.........(1)},$$

$$2(\lambda + \mu) A = -g(\rho' - \rho)\frac{\lambda + 2\mu}{2\mu r}(Arh + B) \qquad \text{.........(2)},$$

provided that $B(rh)^3$ is small compared with A. We know already that in the class of problems we are considering $g\rho$ is small compared with $(\lambda + 2\mu) r$. Now λ and μ are nearly equal, and in these two equations the ratios of the terms in A on the right to those on the left are respectively $\frac{g\rho(\lambda + 2\mu)}{4\mu r(\lambda + \mu)} rh$ and $\frac{g(\rho' - \rho)(\lambda + 2\mu)}{4\mu r(\lambda + \mu)} rh$. In each of these rh is small, by hypothesis, and the other factor is of the order of $\frac{g\rho}{2\mu r}$, which is already known to be small. Hence the terms in A on the right are both products of two small factors, and therefore can be neglected. With this simplification the equations give

$$A = \frac{\rho' - \rho}{2\rho'(\lambda + \mu)} S \qquad \text{.........(3)},$$

$$B = -\frac{2\mu r S}{g\rho'(\lambda + 2\mu)} \qquad \text{.........(4)}.$$

The condition for the validity of the approximation is that

$$(rh)^3 < A/B \qquad \text{.........(5)},$$

while

$$\frac{A}{B} = -\frac{g(\rho' - \rho)(\lambda + 2\mu)}{4\mu(\lambda + \mu)r} \qquad \text{.........(6)}.$$

Taking $g = 1000$, $\lambda = \mu = 5 \times 10^{11}$, and $\rho' - \rho = 1$, all in c.g.s. units, we have

$$\frac{A}{B} = -\frac{7}{10^{10}r} \qquad \text{.........(7)}.$$

Thus our condition gives $\quad r^4h^3 < 7 \times 10^{-10} \qquad$ ....(8),

and if we put h equal to 50 km., or 5×10^6 cm., this gives

$$r < 5 \times 10^{-8}/\text{cm}. \qquad \text{.........(9)}.$$

If the distribution of pressure over the upper surface is simple harmonic, (9) shows that the wave-length must be at least 1200 km. to make the present solution valid. With smaller values of h it may be less.

Substituting in the expressions for the stress-components, and remembering that z cannot exceed h within the crust, we have on rejecting powers of rh above the first, in accordance with the approximation already made,

$$p_{zz} = 2A(\lambda + \mu) \qquad \text{.........(10)},$$

$$p_{xx} = A\lambda\left(1 + \frac{p^2}{r^2}\right) + B\lambda\left(1 - \frac{p^2}{r^2}\right)rz \qquad \text{.........(11)},$$

$$p_{zx} = 0 \qquad \text{.........(12)},$$

$$p_{xy} = \frac{pq}{r^2}\{\lambda A - (\lambda + 2\mu)Brz\} \qquad \text{.........(13)}.$$

If in particular

$$S = \nu \cos \kappa x \qquad \text{.........(14)},$$

we have

$$p_{xy} = 0 \qquad \text{.........(15)},$$

and the greatest stress-difference is

$$|p_{zz} - p_{xx}| = |\{2B\lambda rz - 2A(\lambda + \mu)\}\nu \cos \kappa x|$$

$$= \left(\frac{4\mu\lambda\kappa^2 z}{g\rho'(\lambda + 2\mu)} + \frac{\rho' - \rho}{\rho'}\right)|\nu \cos \kappa x| \qquad \text{.........(16)}.$$

If z has its maximum value h, and the quantities involved have the values already assumed, together with

$$\rho = 2{\cdot}5; \ \rho' = 3{\cdot}5,$$

the first term within the bracket is approximately 3, while the second is 2/7. Thus the greatest stress-difference in a thin crust floating on a fluid will occur at the bottom surface, and will be about 3ν below the places of greatest elevation and depression.

In the thick crust the greatest stress-difference at any depth was found to be $2\nu/e$. Hence the stress-differences in a thin crust may considerably exceed those in a thick one, even when the stress over the original free

surface is the same. If instead the height of the visible inequalities is the same, the extra matter on top must be more on a thin crust because it must make up for the depression of the crust as a whole. On both grounds the internal stress-differences are greater in a thin crust.

10·7. In the foregoing investigation two extreme cases have been considered, the thickness of the crust having been supposed in the one to be great, and in the other to be small, in comparison with the wave-length of the disturbing pressure applied over the surface. In each case the distribution of the stress-differences through the crust has been found. In the case of a thick crust the maximum stress-difference is below the greatest elevations and depressions, and at a depth about $1/2\pi$ of the distance between consecutive ridges. Its amount is about $2\nu/e$, where ν is the maximum pressure applied over the upper surface. When the whole thickness of the crust is less than $1/2\pi$ of the distance between consecutive ridges, the crust bends down as a whole until the extra upward pressure of the fluid on the regions bearing the extra load almost balances the weight of the load. When this is so, the depressed regions are compressed above and stretched below, the opposite holding for the elevated regions. The deformation produces stress-differences that may exceed several times the maximum that can occur in a thick crust of similar materials; the greatest is at the bottom of the crust.

CHAPTER XI

The Theory of Isostasy

"*I* could show you hills, in comparison with which you'd call that a valley."
"No, I shouldn't," said Alice, surprised into contradicting her at last: "a hill *can't* be a valley, you know. That would be nonsense...."
The Red Queen shook her head. "You may call it nonsense if you like," she said, "but *I've* heard nonsense, in comparison with which that would be as sensible as a dictionary!"
LEWIS CARROLL, *Through the Looking-Glass.*

11·1. *The Behaviour of Matter under Shearing Stress.* It was shown in Chapter VIII that the matter of the earth's crust at a depth between 300 km. and 400 km. has had time to cool down since solidification by something of the order of 200°; while the matter at a depth of 700 km. can have cooled by only a few degrees, at the most, from the melting point at the pressure that actually prevails at that depth. We should therefore expect that, while the rocks at the surface possess the properties of ordinary solids, a gradual transition in properties would be associated with increase of depth, and that the matter at a depth of about 700 km. would behave almost as a fluid. Unfortunately, however, this is much too simple a statement of the problem. Several quite different properties are commonly thought characteristic of fluids, but are by no means invariably associated. Thus the use of the term 'fluid' without some preliminary discussion of what is meant by it in the particular context would be certain to lead to confusion. Some account of the properties of matter in its various physical states is therefore necessary.

In studying the development of the earth, especially in relation to its surface features, we are largely concerned with phenomena of change of shape, both temporary and permanent. Hence the physical properties of its constituents that we chiefly need to know will be the relations between the changes of shape that they undergo and the stresses that produce these changes. Other properties of matter, so far as they concern us, will do so only in a subsidiary way. Any classification of substances, to be useful in geophysics, must therefore be based primarily on their behaviour under deforming stress. Such a classification will be outlined in what follows. It does not follow completely any classification known to me, for two reasons. First, no single account I have seen appears to deal adequately with all the properties that require description; and secondly, most accounts are methodologically unsatisfactory in that they use for purposes of definition properties incapable of being tested experimentally.

The dimensions of a body may be altered in two ways. First, it may be compressed or extended by a uniform pressure or tension over the whole

of its surface. This property is called *compressibility*. In many substances, which are called *isotropic*, the dimensions in all directions alter in the same ratio when the pressure is uniform and normal to the surface. If, however, the pressure is not uniform, or if the substance as tested by the above criterion is not isotropic, the dimensions in different directions will in general change in different ratios. Any alteration of size or form in a body can be represented as a combination of changes of volume without change of shape, and changes of shape without change of volume. The simplest type of the latter occurs when a rectangular block has one face clamped, while a tension is applied in the plane of the opposite face. The block will be distorted, its angles ceasing to be right angles, but the volume will remain unaltered. A stress (measured as a force per unit area) that alters angles without altering the volume is called a *shearing stress*. If, again, the tension over each face of the block is normal and uniform, and if the tensions over opposite faces are equal, but those over adjacent faces are different, the block will in general be altered both in volume and in shape, becoming most extended in the direction of the greatest tensions. The angles between lines inclined to the edges of the block will be altered, indicating the presence of shear. If three mutually perpendicular lines meeting in a point in the body remain perpendicular after the deformation, they are called *principal axes of the strain* at that point. In the case of the stress-distribution just considered, lines through any point parallel to the edges of the block are principal axes of the strain. Again, if the stress across a plane at any point is wholly normal to that plane, this plane is called a *principal plane of stress* at that point. At any point there are three mutually perpendicular principal planes of stress, and the stresses across them are called the *principal stresses* at the point. If the substance is isotropic, the principal stresses act along the principal axes of the strain. The difference between the greatest and least of the principal stresses is called the *stress-difference*. It is evident that in an isotropic substance the vanishing of the stress-difference indicates that there is no distortion at the point, the expansion or contraction being the same in all directions. Thus stress-difference and shear in an isotropic substance are always associated.

11·11. *Permanent Set.* Suppose now that a body is deformed in any way, and that the external deforming stresses, after being applied for some time, are removed. Further changes of form will in general follow. The rate of variation of shape may diminish with time in such a way that it is legitimate to infer that it is tending to zero and that the extent of the change of form from the original state is tending to a definite limit. If, for instance, the changes in several successive equal intervals of time decrease like the terms of a decreasing geometrical progression, such an inference will be justified. If the limit is different from the original state, the substance is said to have undergone *permanent set*.

Permanent set follows very different laws in different substances. There may be a limiting stress-difference such that no permanent set is observed unless the stress-difference actually applied exceeds this limit, but such that any greater stress-difference always produces permanent set. This is usually called the *limit of perfect elasticity*, but the term is a bad one, since a body may be strained so as to show no permanent set, while being very far from perfectly elastic in another sense, which will be explained below. It will here be called the *set-point*. It may happen that the set-point is not appreciably different from zero; this is true not only for typical fluids, but for such metals as cast iron. Often, however, it is different from zero.

The nature of the set varies in different cases. It may involve a continuous deformation such as a liquid undergoes, particles originally in contact remaining in contact; or the constituent parts may roll or slip over one another, with or without internal change of form; or parts originally within the body may cease to be in contact with other parts of the body at all. The first type may be called *flow*, the second *elastic hysteresis*, and the last *fracture*; but the property of greatest geophysical interest is the same in all, namely that the substance acquires a permanent elongation in the direction of greatest tension (or least pressure) and a permanent contraction in that of least tension (or greatest pressure); and we are fundamentally concerned with the existence and amount of the set, and not with its nature or the details of the process that brings it about.

The amount of the set may depend not only on the stress-differences applied, but also on how long the stress-difference has exceeded the set-point. It may happen that the set is practically independent of this time; this property is connected with elastic hysteresis. In this case the shape of the body tends to a definite limit under the stress applied, and does not surpass the limit, however long the application continues, unless the stress is further increased.

11·12. *Plasticity and Strength.* When the stress is sufficiently great, however, it will be found that the rate of change of shape shows no sign of falling off when the stress is applied for a long interval. If this is so, the extent of the recovery when the stress is removed is practically independent of the time of application of the stress, so that the set is an increasing function of the time of application, and could theoretically be made to surpass any limit by increasing the time sufficiently. The last property is here called *plasticity*. It is one of the most important properties of matter, since the flow of fluids, the malleability and ductility of some solids, and the brittleness of other solids, are all particular cases of it. The critical stress-difference, above which the rate of change of shape does not decrease when the time of application of the stress increases,

may be called the *strength* of the material; and one substance may be said to be *weaker* than another if it has a smaller strength. Every substance can be made to show plasticity by exposing it to a sufficiently great stress-difference; a body heterogeneous in constitution may, however, show it only when the stress-difference has become great enough to produce it in every constituent*.

The ratio of the stress-difference to twice the rate of shear during plastic flow may be called the *viscosity*. If a body has been undergoing deformation through plasticity, and the external stresses are gradually diminished, the plastic deformation in any element will cease when the stress-difference there sinks to the strength, and thenceforward no further set will be acquired until the stress-difference again surpasses the strength of the material.

11·13. *Rigidity and Elastic Afterworking.* If a body is exposed to a stress-difference insufficient to produce permanent set, and is then released, it may oscillate about its original position, the extent of the oscillations gradually diminishing to zero, or it may return to its original position as a limit without ever passing through it. In either case the tendency to return is said to show *rigidity*. Rigidity and strength are quite distinct properties, but are habitually confused in geological literature. A substance that oscillated about the original position, the oscillations retaining permanently the same mechanical energy, would be called *perfectly elastic*; but such a substance does not exist. The concept of a perfectly elastic body is, however, useful, for in many circumstances the behaviour of real bodies approximates very closely to that of perfectly elastic ones, which therefore serves as a valuable standard of comparison, and is, at the same time, susceptible of exact mathematical treatment. The dying down of the deformation is called *elastic afterworking*. An example of this phenomenon is afforded by certain biscuits containing treacle, notably ginger snaps, when slightly stale. If bent to an extent insufficient to start a crack, and then released, the biscuit may be seen to creep back slowly to its original flat form.

11·14. If a substance that shows rigidity and elastic afterworking when its stress has not exceeded the set-point is afterwards exposed to stress-

* If plastic deformation commences in one constituent, the failure of this constituent to bear the stress-difference falling on it will throw additional stress-difference on the others. An example of this was seen in the problem discussed in 10·6. Hence a plastic yield of the body as a whole will take place more readily than it would if the body were composed entirely of its strongest constituent. This fixes an upper limit to the strength of a heterogeneous body as a whole. The possibility of thus fixing an upper limit requires some emphasis, since it is sometimes thought that the heterogeneity of the earth imposes an insuperable obstacle to the application of any elastic theory. Even if constituents in a state of mechanical mixture become separated, this statement will remain true.

differences greater than the set-point, and is afterwards released, its behaviour during the stress and recovery is a complex process, for rigidity, elastic afterworking and set all take part. It will therefore return only part of the way, if at all, towards its original configuration, and the return will be slow, the oscillations, if any, gradually dying down. To disentangle these effects, and to say how much of the motion is due to each, would require more experimental investigation than has yet been carried out.

11·15. *Definitions of Solids and Fluids.* Typical fluids have no strength and no rigidity. Since elastic afterworking arises only as a property of rigidity, this property also is absent from fluids. Suppose, for instance, that a horizontal plate is suspended in a liquid, and is then moved horizontally through it. The surface remains level, so that gravity does not affect the motion. It is found that, however small the stress acting on the plate may be, it always deforms the fluid, particles originally in vertical columns acquiring horizontal displacements with regard to one another; and that when the plate is no longer acted upon by stress the fluid merely comes to rest, showing no tendency to return. These properties have been defined to be the criteria for absence of strength and rigidity, which are therefore zero in fluids.

But although typical fluids possess neither strength nor rigidity, these properties are not invariably associated. Shoemaker's wax, for instance, is a famous example to the contrary. It is possible to make tuning forks of it, whose free vibrations have a frequency sufficiently high to enable them to give out an audible note; the resilience thus indicated implies rigidity. Yet when one of these forks is left to itself, it gradually flows out under its own weight, until a uniform flat surface has been produced. Hence it has no strength*. In general solids possess both strength and rigidity, but both properties diminish rapidly as the temperature approaches the melting point, and disappear as the substance melts. Strength is usually the first to show great diminution, and in the case of many glasses has quite disappeared some hundreds of degrees below the melting point. Thus to use the term 'fluid,' without carefully specifying whether absence of strength or absence of rigidity is the defining characteristic, would be certain to lead to errors. Each convention would have its advantages, and both have been used for the purpose. Absence of strength has been used as the defining quality by many physical writers. The definition, however, does not appear to have been followed in practice. If we are to have a coherent scheme the definition that should be used should be that actually used in melting-point determinations, which has quite another basis. In such determinations the substance is not kept at a uniform temperature for as long a time as would be required, for instance,

* Lord Kelvin, *Baltimore Lectures*, Camb. Univ. Press, 1904, 9–10.

to make pitch at ordinary temperatures acquire a flat surface. Thus whatever the practical criterion of the fluid state may be, it is not absence of strength. Strength, indeed, is lost by all impure substances, and by many pure ones, at temperatures far below the accepted melting points. The property actually used is the acquirement of mobility; that is, the substance is considered to be fluid when it can be poured. Now if one tries to pour a substance possessing rigidity, the flow is resisted by rigidity until the stress-difference surpasses the strength, and afterwards by the resistance to plastic flow, so long as this continues; thus mobility implies the smallness of both these qualities. The absence of rigidity in the liquid state is again shown by the absence of any tendency towards elastic recovery of form when the substance has been poured. It therefore appears that the characteristic property of fluids is twofold: they have zero rigidities, and their viscosities are small in comparison with those of the same substances in the solid state. Either of these properties might be used as an expression of the practical criterion, the great reduction in viscosity on melting being probably the more convenient; but they are closely associated in actual substances, and it will be a matter for no surprise if rigidity in a substance is found to be present almost or quite up to the melting point. We shall therefore recognize the fluid state of a substance by the absence of rigidity and by the smallness of the viscosity in comparison with that in other states of the same material. Other states characterized by high viscosity, with or without rigidity and strength, will be called solid. Thus pitch at ordinary temperatures will be regarded as a solid.

11·16. *Properties of Solids.* It will be seen that this definition of the solid state suggests a further classification of solids according to their possession of rigidity and strength. It is possible, though not certain, that all solids possess rigidity. They may, however, be devoid of strength.

There is only one state of solids in which they are quite lacking in strength. In this state they are amorphous and practically uniform throughout. There is another state, however, bearing, at first sight, a close resemblance to the last, in so far as solids in this state are also amorphous and uniform, but differing from it in the possession of considerable strength. Both states are commonly described as *vitreous* or *glassy*. The difference between them is so important, however, as to merit a difference in nomenclature. The former will here be called the *liquevitreous* and the latter the *durovitreous* state. Any given vitreous substance is of the former type above a certain critical temperature, and of the second type below that temperature. This critical temperature is, of course, to be carefully distinguished from the critical temperature encountered in the discussion of the transition between the liquid and gaseous states. This is true for instance of all ordinary kinds of glass and of fused silica. The transition is usually gradual. Thus in one kind of hard glass with

a critical temperature of 750° the strength below 750° has been found to vary roughly as $(750 - V)^2$, where V is the temperature. In a soft glass the critical temperature may be something like 400°–450°. All vitreous substances are isotropic.

Another type of solid state is the state characterized by the existence of a definite crystalline form. The crystalline state is inherently weaker than the durovitreous, because every crystal possesses cleavage planes, over which slip may take place for stress-differences much less than are required to produce plastic flow in the durovitreous state. Dr A. A. Griffith informs me, for instance, that he has maintained vitreous silica at room temperature for a week at an elastic extension of 5 per cent., without detecting any flow. If there had been a flow of 0·005 per cent. of the length of the fibre he could have detected it. This corresponds to a far greater stress-difference than could be withstood for so long by any substance in the crystalline state.

No crystals are isotropic; even cubic ones behave differently with regard to stresses along the diagonals and normal to the faces. In actual rocks, however, the crystals are orientated in all directions and usually are not even all of one kind. Hence the differences in elastic properties in different directions shown by different crystals will be expected to balance one another, and so long as we are dealing with masses of rock so large as to include a very large number of crystals, we may regard them as isotropic.

Vitreous substances often tend to crystallize when near particular temperatures, which may be above or below the critical temperature. If the temperature of crystallization is above the critical temperature, crystallization is resisted only by viscosity, and therefore must occur if sufficient time is allowed. If, however, it is below the critical temperature, the strength of the durovitreous matter may be enough to withstand the stresses involved in crystallization, and crystallization will then be impossible. The strength increases rapidly with decrease of temperature; hence if a substance is cooled sufficiently quickly through the liquevitreous state and the hotter part of the durovitreous state, crystallization may not have time to start, and may be afterwards permanently prevented.

The lack of strength of liquevitreous substances, combined with the use of the same name for both them and durovitreous substances, has led to a widespread belief that all substances of both classes are 'supercooled liquids' and able to flow to an indefinite extent under any stress-difference, however small, provided it is maintained for a sufficient time. The latter statement is true of liquevitreous substances, but not of durovitreous ones; the former depends on the definition of 'supercooling,' and will not arise in the present work. Ordinary window glass resists the stress due to its own weight for hundreds of years without appreciable set, a longer time than has been available in any

experiments supporting the statement that all glasses are devoid of strength*.

In addition to the above types of solid, there are two types of fragmentary solid. Some solids are composed of a vast number of small particles, which may be fastened together by a matrix or quite free. In the former case the solid follows the rules for mixtures already elaborated. The only need for special warning is that the matrix may be stronger than the particles; as may be seen from an inspection of an ordinary piece of broken brick concrete, where the fracture goes through the fragments of brick instead of through the cement or along the boundaries between brick and cement. If there is no matrix, strain again follows the usual rules until it becomes great enough to make the particles roll or slide over one another, after which the substance behaves more or less like a solid showing elastic hysteresis and plasticity together.

11·2. *Probable Mechanical Properties of the Earth's Crust.* Coming now to the application to the earth of this account of the properties of matter, we notice that rigidity, viscosity, and strength all increase as the temperature falls from the melting point, but that whereas strength may remain zero until cooling through some hundreds of degrees has taken place, viscosity is certainly, and rigidity probably, considerable as soon as the substance is below the melting point. We shall expect that the rocks of the earth's crust, where they have cooled by only a few degrees since solidification, will have no strength, but that where they have cooled by some hundreds of degrees their strength may become considerable. The question is complicated by the high pressures prevailing at considerable depths in the earth's crust; but it will be observed that by hypothesis the initial temperature at any depth was the melting point appropriate to the pressure at that depth, and that just below the melting point every substance is lacking in strength, but has by definition considerable viscosity and probably considerable rigidity. These statements are true whatever the pressure; all we need to assume is that the temperature interval between solidification and the acquirement of strength remains of the same order of magnitude at the pressures existing at depths of some hundreds of kilometres as it is at the surface. If this plausible hypothesis be granted, it follows that the rocks at a depth of 700 km. can have no appreciable strength, that those at a depth of 300 km. may be just acquiring it, and that those at a depth of 30 km. may be very strong. Rocks at all depths, however, may have great rigidity. The rigidity may be

* For other considerations concerning the relations of the vitreous states to the crystalline and liquid states, see Beilby, *Aggregation and Flow of Solids*, 1921; A. A. Griffith, *Phil. Trans.* A, **221**, 1920, 163–198; Jeffreys, *Proc. Camb. Phil. Soc.* **24**, 1928, 19–31. My own view is that the fundamental states of matter should be considered as gaseous, crystalline, and vitreous. The liquid state is a compromise between the gaseous and vitreous ones; the fact that it is a common state tends to convey an entirely erroneous impression that it is easy to understand.

greater than that of rocks of the same constitution at the surface; for the rigidity is raised by pressure, and this effect may be enough to counteract the reduction due to the greater temperature.

11·21. Let us consider the stresses in the crust due to the weight of a series of parallel ranges of mountains. We saw in 10·5 that if the elastic constants (i.e. the incompressibility and the rigidity) are the same throughout the crust, and if the depth of the crust is more than, say, half the distance between consecutive ridges, the maximum stress-difference will occur at a depth approximately equal to $1/2\pi$ of the distance between consecutive ridges, and will be equal to $2/e$ times the weight of the excess load per unit area where it is greatest. Now consider a series of ranges 3 km. in height above the mean level of the land. Taking the density as 2·7, we find the maximum stress-difference to be 6×10^8 dynes/cm.^{2} The crushing strength of basalt is $1{\cdot}2 \times 10^9$ dynes/cm.^{2} Thus if the earth's crust were uniform and of the same strength as basalt in the laboratory, the weight of mountain ranges comparable with the Rockies and the Alps would be insufficient to produce plastic deformation of the crust at any depth; the weight of the Himalayas might be just sufficient.

11·22. When the reduction of strength with depth is taken into account, this statement evidently requires some qualification. If the earth has a thin strong crust, resting on a deep layer with no strength, any inequality capable of producing appreciable stress-differences in a uniform crust must produce plastic deformation in the actual crust. The weak matter will flow out laterally, and the upper crust will bend down, the amount of the depression being determined by the elasticity of the crust and the density of the weak matter, in accordance with the discussion of Chapter X. In the process the stress-differences in the upper crust will be increased, becoming greater than the maximum stress-difference in a uniform crust. If the distance between consecutive ridges is large in comparison with the depth of the layer of weakness, the approximation of 10·6 will become applicable. In this case the extent of the depression is such that the weight per unit area of the weak matter that flows out is nearly equal to the pressure applied over the surface, which is itself equal to the weight of added matter per unit area. Thus the effect of flow in a thin crust will be that the total quantity of matter in a column of given cross section down to a place in the layer of weakness to a given depth below the original surface will be unaltered by surface displacements.

11·23. The discussion of the last paragraph is still too simple an account of the facts. In the actual crust we should not expect that the rocks of the crust would have a uniform strength down to a certain depth, and no strength below that depth. Thus the last result does not represent the true state of affairs, but it provides a basis for further investigation. We

should expect a gradual variation of strength, possibly distributed over some hundreds of kilometres in depth; for the extent of cooling since solidification decreases continuously with depth, while the pressure increases continuously. Even in glasses of uniform composition the reduction of strength in passing from the durovitreous state to the liquevitreous state is gradual, and in the earth's crust the change must be gradual both for this reason and because rocks are in general mixtures. Hence a continuous decrease of strength would be expected, starting at the layer of greatest strength, whose depth is so far unknown, and finishing at a depth of about 400 km., below which strength is probably absent. The consequence will be that additional pressure applied over the surface will produce plastic deformation, not only in the region of zero strength, but also in all places above it where the stress-differences produced exceed the strength. Thus it is possible that the major part of the flow involved in the adjustment of the earth's crust to superficial inequalities may take place at depths much less than 400 km. If so, the crust may be treated as thin when the horizontal extent of the inequalities considered is much less than the 3300 km. given by 10·6 (8) for a crust of this thickness.

11·24. It will be noticed that the uniformity of mass per unit area over the earth's surface, inferred to hold when the crust is thin, would imply the absence of any superficial inequality if the density of the added matter was equal to that of the matter where flow first occurs; for to produce this uniformity the depth of matter that flows out would have to be equal to that of the added matter, and thus the depth of the matter in a vertical column would be the same as before. Hence no superficial inequality could persist if the crust were thin and the density uniform. If, however, the lightest matter was on top, when the adjustment was complete the depth of matter that flowed out would be less than that of the added matter, and then a projection would remain on the surface; but the greater part of the added matter would sink below the original surface in the process, and the projecting portion would correspond only to the visible portion of an iceberg, the part below the surface corresponding to the much larger portion of the iceberg that lies below the surface of the water.

11·25. The discussion of Chapter X referred to a flat earth of infinite extent. The results are readily adapted to give an interpretation that takes account of the finite size of the earth. It was seen that the stress-differences decreased rapidly with depth, so that a depth could always be found such that the stress-differences below it produced by a given load on the surface would not exceed any assigned limit. The hydrostatic state, in which the stress-differences are zero, would be approximately undisturbed below that depth. Now in the actual earth, if the strength below a certain depth is zero, the hydrostatic state must exist below that depth and be unaffected by surface load. An important property of the hydrostatic

state is that the surfaces of equal density, equal gravitation potential, and equal pressure coincide. Our condition is therefore that the pressure, density, and gravitation potential below the layer of weakness are unaffected by surface load; and if the load is of such an extent that the crust can be regarded as thin, the mass per unit cross section in a column, cut down to some definite equipotential surface within the layer of weakness, will be unaffected by surface load, and will be the same for all places.

11·26. It has been seen that a great difference is to be expected between the adjustment of the earth's crust, on the one hand, to inequalities whose horizontal extent is so small that the stress-differences they produce reach their maxima far above the layer of weakness, and on the other hand to inequalities whose horizontal extent is large. In the former case the flow produced is inappreciable, and hardly affects the mass in a vertical column; any matter added, or any valley denuded away, produces directly a disturbance of gravity, which is not appreciably reduced by outflow below. In the latter case, the addition of extra mass to the surface will produce a disturbance to gravity in its neighbourhood, but the reduction of mass below will cause the anomaly to be much less than if the earth was undeformable. A great deal of observational evidence has been acquired concerning the effect of surface inequalities on gravity, and this will now enable quantitative tests to be applied to the theory so far developed.

11·3. *Effect of Uncompensated Surface Inequalities on Gravity.* The older geologists and geodesists regarded mountains as composed of matter of much the same density as the rest of the crust, and it was not realized that their weight would be expected to produce any deformation of the material below them, nor that the density of the matter below a mountain range might differ systematically from that of the matter at an equal depth below a plain or even an ocean. Now if a mountain is considered merely as an extra mass superposed on a previously uniform crust, and its deforming effect on the interior is ignored, it is possible to compute its gravitational attraction on bodies in its neighbourhood. The attraction can also be found experimentally, and the result compared with that calculated. The experiment was carried out on several mountains during the eighteenth century, but the results were in conflict.

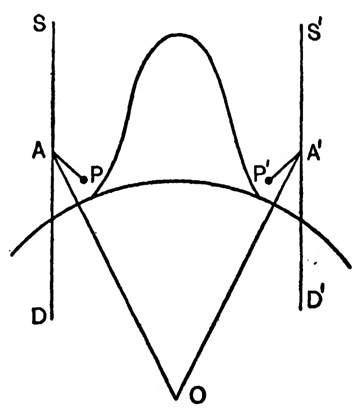


Fig. 14.

The principle of the measurement of the attraction of a mountain may be illustrated by means of the above figure. Let A and A' be two pivots on opposite sides of the mountain, from which two pendulums are suspended. Let O be the centre of the earth, supposed spherical. Let AS,

$A'S'$ be lines joining A and A' to a fixed star in the plane AOA', and let AD, $A'D'$ be their prolongations past A and A'. Then the angle AOA' is nearly AA'/R, where R is the radius of the earth. Now AA' can be found from surveying operations, and therefore AOA' is determinable. Let its value be α. But we have, since AS and $A'S'$ are parallel,

$$OAD + OA'D' = AOA' = \alpha.$$

Next, the zenith at A, say, is defined to be on the prolongation of the plumb-line past its point of support. Hence the zenith distance of the star is equal to PAD. But the zenith distance is observable. Thus we can find PAD and $P'A'D'$ from observation. Let their sum be β. Then we have by subtraction

$$OAP + OA'P' = \beta - \alpha.$$

But if the mountain had no deflecting effect, the two pendulums would point straight towards the centre of the earth, and therefore OAP and $OA'P'$ would both be zero. Hence $\beta - \alpha$ is equal to the sum of the deflexions of the plumb-line at the two stations owing to the attraction of the mountain.

Now if g be the intensity of gravity, and the bob of the pendulum AP be exposed to a small horizontal acceleration γ due to the mountain, the deflexion of the pendulum is, by the ordinary rules of statics, γ/g. Hence if γ' refer to the other pendulum,

$$\gamma + \gamma' = g\,(\beta - \alpha).$$

The accelerations γ and γ' are calculable from the law of gravitation, if the region has been surveyed, and therefore this relation affords a test of the theory.

The first attempt to use this method was apparently made by Bouguer*. The sum $\beta - \alpha$ of the deflexions due to the mountain Chimborazo was found to be very much less than that calculated from the law of gravitation. At that time the constant of gravitation was known only vaguely, but the deflexion of the plumb-line was much less than could be reconciled with any reasonable value. Maskelyne, in 1774, repeated the experiment at Schiehallion, in Perthshire. His results gave

$$\alpha = 41''; \quad \beta = 53''.$$

The value of $\gamma + \gamma'$, namely

$$\gamma + \gamma' = 0{\cdot}06 \text{ cm./sec.}^2$$

inferred from this, was used to estimate the constant of gravitation, and hence the mean density of the earth. The density was found to be 4·71. This is decidedly lower than modern estimates, but the discrepancy is much less than that found by Bouguer. A repetition by Petit† in the Pyrenees showed that their attraction was not only small, but actually

* *La Figure de la Terre*, Paris, 1749.

† *Comptes Rendus*, **29**, 1849, 729–734.

negative; the plumb-line appeared to be deflected away from the mountains. Indeed the attraction of mountains was generally found to be nearer to zero than to the values calculated on the supposition that the underlying matter was of normal density. Schiehallion was an exception to the general rule; the reason it was exceptional was probably that it was of smaller size, so that the attraction of the displaced matter below was mainly vertical and did not deflect the pendulum much.

The modern development began with a discussion by Archdeacon Pratt* of the deflexions of gravity observed in India by Everest. These deflexions were too great to be attributed to errors of observation, but they were only about a third of what would be produced by the attraction of the mountains, treated as additional masses.

11·31. Another method of testing the attraction of a mountain is to consider the intensity of gravity on top of it, instead of the direction at its sides. For this purpose it will be sufficient to regard the mountain as flat; this is justifiable, for the width of a mountain or range of mountains is always much greater than its height. If the mass of the earth be M, its mean density ρ_0 and its radius a, the intensity of gravity at the surface is

$$g = fM/a^2 = \tfrac{4}{3}\pi f\rho_0 a \qquad \text{.........(1)},$$

where f is the constant of gravitation. At a height h above the surface, in the free air, the intensity of gravity is

$$\frac{fM}{(a+h)^2} = \frac{fM}{a^2}\left(1 - \frac{2h}{a}\right) = g\left(1 - \frac{2h}{a}\right) \qquad \text{.........(2)},$$

h^2 being neglected. If instead of the interval between sea-level and height h being filled with air, it is occupied by a mountain of density ρ', it is known from the theory of attractions that it will add an amount $2\pi f\rho' h$ to the attraction above it. Hence the total intensity of gravity at the top of the mountain, on the hypothesis again that the matter below it is of normal density, is

$$g\left(1 - \frac{2h}{a}\right) + 2\pi f\rho' h = g\left(1 - \frac{2h}{a} + \frac{2\pi f\rho' h}{g}\right)$$
$$= g\left(1 - \frac{2h}{a} + \frac{3}{2}\frac{\rho'}{\rho_0}\frac{h}{a}\right) \qquad \text{.........(3)}$$

by (1). Hence the excess of gravity at the top of a mountain over its value at sea-level is $-\dfrac{2gh}{a}\left(1 - \dfrac{3}{4}\dfrac{\rho'}{\rho_0}\right)$. In actual cases ρ' is always less than ρ_0, and therefore this anomaly is always negative. In other words, the gravity on a mountain top is less than elsewhere. This formula was obtained by Bouguer. As in the case of the deflexion of the pendulum, it was found when tested to give considerable errors; the actual anomaly of

* *Phil. Trans.* **145**, 1855, 53–100.

gravity on the top of a mountain was nearer to $-2gh/a$ than to the Bouguer formula.

11·4. *Compensation.* It is seen from an examination of the formula for the gravity anomaly that only the second term arises from the attraction of the mountain itself. Thus the statement that the gravity anomaly on the top of a mountain is equal to $-2gh/a$ implies that it is the same as if the mass of the mountain were zero, its height remaining the same. The hypothesis that the mountain is an egg-shell gives better accordance with truth than the hypothesis that its weight does not deform the interior. A hypothesis virtually identical with the former was offered by Boscovich* when he suggested that mountains were swellings caused by the earth's internal heat, no extra matter being added. Cavendish† also suggested, in a manuscript of 1772–4, that the matter below the stations in the Andes was of abnormally low density. The use of this formula was reintroduced by Faye‡; it is usually known as the Free Air hypothesis.

Thus with regard both to the deflexion of the plumb-line and to the int nsity of gravity, the attraction of a mountain is very much less than it ıld be if a deformation were not produced below. But this is exactly w' would be expected from the theory of the deformation of the layer of ess by superficial load. It has been seen that the mass under a given st in any deformation of a thin crust lying on a weak interior is nearly to that added on top. The two effects balance, so that the mass ertical column of unit cross section is unaltered, instead of being sed by the mass per unit area of the mountain added, or decreased mass per unit area denuded away from a valley. Hence the extra tion in any direction due to a mountain is partly neutralized by the ution in the attraction due to the loss of mass below it. The failure Bouguer formula to fit the facts, and the rough agreement with the Free Air formula, are therefore in accordance with the theory elaborated of the strength of the earth's interior.

11. A qualitative explanation of the facts discovered by Pratt was oy Airy§. His theory is practically that developed here; he con- l the earth as having a thin solid crust, supported on a weak, but cessarily fluid, substratum, and showed that the elastic resistance crust to bending might be so small that any extra load added would l he crust down until the load was balanced almost wholly by the u l pressure of the material below. The smallness of the disturbance of rity due to the visible surface topography follows as a natural coı .uence. Airy remarks indeed that Pratt's results ought to have been

* *De Litteraria Expeditione per Pontificiam Ditionem*, 1750, p. 475; or Todhunter, *Mathematical Theories of Attraction*, 1, 313.

† *Papers*, 2, 1921, 404.

‡ *Comptes Rendus*, 90, 1880, 1443–1446.

§ *Phil. Trans.* 145, 1855, 101–104.

anticipated, for the notion of a crustal layer with a magmatic ıbstratum was a familiar one at the time. The earliest reference to it tha the writer has traced was by Sir John Herschel*. It was inferred from tl fact that the vertical variation of temperature in the crust, if it conti ued to a depth of some tens of kilometres, would lead to fusion there; aı d though the argument is much too simple a statement of the problem, the estimate it gives of the thickness of the strong layer is not far different fı om that in 11·2. Airy, in his paper, considered thicknesses up to 160 km.

Pratt did not accept Airy's theory, on three grounds, which a re now of historic interest only†. It has been noticed in 11·24 that if the under layer was of the same density as the upper one when adjustment took place, balance would not be attained until the surface inequalities were quite effaced. If the lower layer was the less dense, equilibrium could n ot be attained at all until our approximations had broken down. Pratt, consid ering the lower layer to be of the same material as the upper one, but liquid ins tead of solid, considered that on Airy's theory the lower layer would be the less dense, and that the mechanism would fail. But now that the lower lay r is known to be of a different material from the upper one, with a greater d ty, this objection has lost its force. It is not, however, obvious tha ak layer reaches up to the top of the lower layer of seismology; it m ɔe at a considerable depth within it. Its depth can be determined, if only by examining its effects on gravity.

Pratt proposed the alternative hypothesis that the developm surface features is due to the vertical expansion of columns of rock to a certain uniform depth; the expansion is the same at all points same column, but differs from one column to another. There is no c of mass within any column, and therefore the smallness of the distuı of gravity agrees as well with this theory as with Airy's. Ther however, fatal objections to it, for it does not account for the facts very cases where they are best ascertained. Geological observati established that mountain formation is not a matter of simple upl that folding and thrusting take place to such an extent that the r a transverse section of a range must have been originally spread that their horizontal width was greater than now by tens or even hu of kilometres. Denudation has lowered the surface, especially in tainous regions, by distances comparable with the present heights mountains; and redeposition in deltas has raised the surface th comparable amounts. All these phenomena are essentially horizon placements of mass. Observations of gravity make it clear that t not lead to great departures from isostasy, yet by their very nature hypothesis is inapplicable to explain how compensation is maintained.

Modifications of the hypothesis to meet these difficulties have b

* *Proc. Geol. Soc.* **2**, 1837, 597.

† *Phil. Trans.* **149**, 1859, 745–778. Cf. also *Geol. Mag.* **65**, 1928, 280.

proposed, but each seems to introduce new ones. The abandonment of the restriction to vertical movement makes it possible to explain folding, in certain conditions. But if the theory is to do so the rock whose initial expansion starts the movement must become stronger than the rocks folded; otherwise it will itself yield and be forced vertically upwards. The conditions associated with expansion are not often associated in nature with an increase of strength. The folding, again, will not take place on the spot, but some distance away among the rocks not yet altered, and will therefore involve horizontal compression and increase of mass per unit area, which will have to be compensated by further horizontal movements below. The hypothesis therefore still fails to explain the compensation of mountains.

It has been suggested also that contraction of volume may lower the surface and thereby create a place where deposition will occur. This only introduces a further dilemma. Either the contraction involves horizontal movement, in which case it will produce a departure from isostasy which will persist until late in the process of deposition: or the movement will be purely vertical, and accumulation of sediments will destroy isostasy.

Pratt's hypothesis therefore fails to explain how isostasy is maintained in any of the principal geological processes leading to the development of surface features. The alternative hypothesis that surface changes lead to changes of density of just such an amount as to maintain isostasy succeeds no better. Changes of mass cannot be admitted, and the maintenance of isostasy during the formation of a delta requires that the density should diminish as the pressure increases, contrary to the properties of ordinary matter.

The theory fares no better in accounting for the compensation of the ocean basins. Whereas the rocks of the continents are mainly granitic or derived from granite, those of oceanic islands, which are the best approximation to the ocean floor yet accessible to us, are basic. The difference between the continents and the ocean floor is a fundamental one of chemical composition, and could not have been produced by a mere change of physical state.

It is not intended, of course, to imply that expansions and contractions, raising and lowering the solid surface, have not occurred, or that they would necessarily destroy isostasy if they did occur. What does seem clear is that they cannot account for compensation in just those most conspicuous features of surface relief where it is best known to exist.

In what follows the mechanics of Airy's theory, which accounts automatically for the compensation of all surface irregularities, however produced, will be adopted. It provides that every surface projection or depression large enough to produce permanent deformation in the weak layer will be compensated. The strong layer will be usually called the *lithosphere*, and the underlying one the *asthenosphere*, following Barrell*.

* *Journal of Geology*, 1914–15.

The distinction between them is not very definite, as the actual transition in properties is probably continuous, but it is nevertheless worth making. The thickness of the lithosphere is a matter for special investigation; it is not obviously identical with any of the layers revealed by seismology or even with any set of them.

It may be remarked that Airy has never received proper credit for his contribution to the theory; in fact it is customary for geodetic writers, even when using his principle of the squeezing out of the weak but dense substratum, to attribute the essentials of the theory to Pratt, whose reputation can rest quite well on his discovery of the facts.

The name *isostasy* was coined by Major C. E. Dutton*. Some writers use *compensation* to denote the fact of approximate uniformity of mass over the earth within vertical columns of the same cross section extending down to a standard equipotential surface, and restrict 'isostasy' to the physical process that leads to the establishment of this state.

11·42. The depression of the crust by the weight of the extra matter on top must in general extend some distance beyond the margin of that matter, so that the compensation is not strictly confined to places vertically below the surface inequalities; it is spread some distance around them. This is expressed by saying that the compensation is *regional*. If the compensation was not spread out in this way it would be *local*. Also the compensation may be uniformly distributed with regard to depth, or concentrated in one or more definite layers. In most of the published work relating to isostasy the idea of uniform compensation is attributed to Pratt, since he adopted a uniform expansion in his explanation of surface inequalities, and compensation concentrated in a single layer to Airy. This difference is much less fundamental than the one that actually led Pratt to reject Airy's interpretation of the data. It is easy to see, in fact, that uniform compensation would be entirely consistent with Airy's mechanism. Suppose that the density of additional matter on top is ρ', and its thickness k. Then the additional mass per unit area is $\rho' k$. Again, let the density of the matter at the level where outflow takes place be ρ_0. Then compensation would be attained if the depression of the crust was $\rho' k/\rho_0$, for this would make the mass expelled below equal to that added on top. If we neglect compressibility, this depression requires that the density after deformation at a height x above a fixed equipotential surface within the layer of weakness must be equal to the original density at height $x + \rho' k/\rho_0$, and is therefore $\rho + \dfrac{\rho' k}{\rho_0}\dfrac{d\rho}{dx}$, where ρ is the original density at height x. The effect of the deformation is to increase the density at a given place by $\dfrac{\rho' k}{\rho_0}\dfrac{d\rho}{dx}$, which is in general negative, and is constant

* *Bull. Wash. Phil. Soc.* **11**, 1889. 51–64; reprinted *J. Wash. Acad. Sci.* **15**, 1925, 359–369.

if ρ is a linear function of x. Uniform compensation down to a given depth is therefore to be expected if the variation of density down to that level is linear*, but if the density changes suddenly at any level the deficiency of mass will appear to be concentrated near that level. The decision between uniform compensation and compensation at a single depth does not depend on the mechanism of isostasy, but only on the normal distribution of density with depth. Now the evidence of near earthquakes makes it clear that the variation of density is actually concentrated near particular levels, namely the transitions between the granitic, intermediate, and lower layers. The transitions may be gradual, but the changes in properties within each of the upper layers are small compared with those from one layer to the next. Compensation must therefore actually be concentrated near the bases of the granitic and intermediate layers.

11·5. Let us now consider the question from the other side and see what observation of gravity tells us, and what it is capable of telling us. Geodesists have proceeded as a rule by comparing the observed intensity and direction of gravity with those calculated on three hypotheses: the Bouguer hypothesis, the Free Air hypothesis, and that of uniform local compensation distributed to various depths. The effects of compensation are easiest to calculate if it is uniform and local, but compensation concentrated at the base of a single uniform layer has recently been examined in an important work by Heiskanen†. The largest investigation of gravity carried out so far is that of the United States Coast and Geodetic Survey, under Hayford and Bowie. Hayford, following up some earlier work of Putnam, studied the deflexions of gravity within the United States, and found that the uniform compensation that suited them best extended to a depth of 122 km.‡ Bowie later considered the anomalies in the intensity of gravity, and found the best depth to be 96 km.§ To indicate the degree of agreement the results in the tables on p. 196 are reproduced from his paper. In this table H is the depth reached by the compensation, assumed uniform.

It will be seen that if H was infinite, the change of density would be zero at all depths, and there would be no disturbance of gravity other than that due to the elevation itself. Thus the case of infinite depth of compensation corresponds to the hypothesis of an undeformable earth. If H is zero, the compensation is all concentrated in the surface, so that the mountain has no attraction, and we have the Free Air hypothesis. Thus the hypothesis of uniform compensation includes the Bouguer and Free Air hypotheses as particular cases.

* Jeffreys, *Proc. Roy. Soc.* A, **100**, 1922, 131.

† *Veröff. d. Finn. Geodät. Inst.* 1924, 1–96.

‡ *Figure of the Earth and Isostasy*, 1909; *Supplementary Investigation in 1909*, 1910.

§ *Investigations of Gravity and Isostasy*, U.S. Coast and Geodetic Survey, Special Publication **40**, 1917.

In the first place, we see by inspection of these tables that no solution makes the mean of all the residuals without regard to sign less than 0·018 cm./sec.² This is to be regarded as representing the irregular variation, and no mean residual can be considered significant unless it decidedly exceeds this standard. We notice then that no residual for coastal stations or stations within 325 km. of the coast can give useful information concerning the accuracies of the four hypotheses. In interior stations, not in mountainous regions, the Bouguer hypothesis begins to fail, giving a mean residual of 0·033 without, and – 0·028 with, regard to sign. In the mountainous regions, below the general level, the mean Bouguer anomaly has become 0·108 without, and – 0·107 with, regard to sign; and in mountainous regions, above the general level, it reaches 0·111 without, and

Mean residuals without regard to sign.

	$H=\infty$	$H=0$	$H=114$ km.	$H=60$ km.
Coast stations	0·021	0·022	0·018	0·012
Stations near coast (within 325 km.)	0·025	0·023	0·021	0·020
Stations in interior (not in mountainous regions)	0·033	0·020	0·019	0·019
Stations in mountainous regions, below general level	0·108	0·024	0·020	0·018
Stations in mountainous regions, above general level	0·111	0·059	0·017	0·022

Mean residuals with regard to sign.

	$H=\infty$	$H=0$	$H=114$ km.	$H=60$ km.
Coast stations	+ 0·017	+ 0·017	– 0·009	– 0·003
Stations near coast (within 325 km.)	+ 0·004	+ 0·017	– 0·001	+ 0·002
Stations in interior (not in mountainous regions)	– 0·028	+ 0·009	– 0·001	– 0·001
Stations in mountainous regions, below general level	– 0·107	– 0·008	– 0·003	0·000
Stations in mountainous regions, above general level	– 0·110	+ 0·058	+ 0·001	+ 0·016

The unit in each case is 1 cm./sec.² (not 1 dyne as stated by the original author).

– 0·110 with, regard to sign. In no case has the anomaly on either Hayford hypothesis risen above the ordinary limits of irregular variation. This is sufficient to make the Bouguer hypothesis untenable; the close agreement in absolute value between the residuals with and without regard to sign shows that the hypothesis is in error in the same sense at nearly every station. Some form of compensation must therefore be admitted.

The Free Air hypothesis is in sufficient agreement with the facts in the first four lines of each table, but breaks down completely when applied to the mountain stations, above the general level of the neighbourhood. As on the Bouguer hypothesis, the residual has the same sign at nearly every station; but the error is systematically in the opposite direction.

The discordance, in opposite directions, of the hypotheses of the undeformable earth and of massless mountains with the observed variation of gravitation is conclusive evidence of the existence of some form of

compensation within a finite depth. It is found that the residuals are made smallest for mountainous regions if the depth of compensation is taken to be 96 km., and for the less elevated regions if this depth is about 60 km.; but where the inequalities are small they could produce little disturbance of gravity even if they were uncompensated, and therefore little weight can be attached to a determination of the depth of compensation from them. The data are quite well represented by a uniform depth of compensation of 90 to 100 km. Helmert's estimate from European data was about 118 km.

Heiskanen has compared the gravity observations in the United States with those calculated on various hypotheses of compensation concentrated at the base of a single layer. He finds that the average depth of the bottom of this layer that fits the facts best is about 50 km., and that the fit it gives is slightly better than that given by the hypothesis of uniform compensation at its best. He has extended the comparison to the Swiss Alps, the Austrian Alps, the Harz Mountains, and the Caucasus, and in each case he finds that compensation concentrated at a single level fits the observations as well as or rather better than compensation uniformly distributed down to a certain depth. His result, so far as it goes, confirms our inference from other considerations that the compensation should be concentrated in one or two surfaces. The single surface he obtains is estimated to be at a depth below sea level of 50 km. in the United States, 77 km. in the Caucasus, and 41 km. in the Alps. Since the hypothesis he is testing refers explicitly to an upper layer of variable thickness, the actual thickness in the mountains exceeds the normal thickness appreciably.

11·6. Heiskanen's discussion, like those of Helmert, Hayford and Bowie, has the valuable feature that it rests on a definite assumption with a unique answer, which can be made the basis of further work. It does not follow that the answer is in close agreement with the facts till certain other questions have been examined, namely,

(1) whether the observations would be equally well fitted by several different distributions of density in the crust;

(2) if so, which of these distributions is favoured by other evidence;

(3) whether the differences between theory and observation that still remain are too great to be attributed to errors of observation, and if so, to interpret them.

11·61. The first question is capable of a definite answer. It is only some distributions of gravity over a sphere that can be explained at all by internal anomalies of density: the condition for such explanation is practically that when the values of gravity at the surface are expressed as the sum of an infinite series of spherical harmonics, the series shall converge like a geometrical progression. When an arbitrary function is so expressed the series usually converges only like Σn^{-p}, where n is the number

of the term and p a small positive number. Special conditions are therefore necessary for any solution to exist. Further, the convergence fixes a limit to the depth where the anomalies of density are situated. If they are at depth h, and the radius of the earth is a, the series will converge faster than Σx^n, where*

$$x = e^{-h/a} \qquad \text{.........(1).}$$

If then x is determined from the series,

$$h < a \log \frac{1}{x} \qquad \text{.........(2).}$$

If the wave-length of the gravity anomaly to be explained is λ, measured as twice the distance from a maximum to the nearest minimum, there is a strong presumption that it arises from an anomaly of density at a depth of λ/π at most.

On the other hand, if the anomalies of gravity can be explained by inequalities of density at any depth, they can be explained equally well by such inequalities at any smaller depth. In fact if we have a particle of unit mass at distance $a - h$ from the centre of the sphere, and we take any concentric sphere of radius b between $a - h$ and a, the field produced by the particle at points outside the sphere of radius b is exactly equivalent to that produced by a distribution of matter over this sphere of density $\dfrac{b^2 - (a - h)^2}{4\pi b r^3}$ per unit area, where r is the distance from the particle to the point on the sphere of radius b†. But if b was less than $a - h$, no distribution of matter over the sphere would give an external field equivalent to that of the particle.

Now we can consider the disturbance of gravity due to the visible inequalities as calculated and subtracted from observed gravity, leaving the Bouguer anomalies as the phenomenon to be explained. What Heiskanen has shown is that the greater part of these can be explained by a surface density at a determinate depth, of amount closely correlated with the surface height. The fit becomes definitely worse if the depth is increased; but the data will be represented precisely as well by a surface density at any smaller depth, provided the extra mass under any spot of the surface postulated in Heiskanen's solution is spread out around it according to the rule just given. This amounts to saying that the visible protuberances and deficiencies may either be compensated purely locally, as in Heiskanen's discussion, or regionally at a smaller depth. Gravity observations cannot decide between these possibilities‡, and the depths found by Heiskanen are upper limits to the depths where the compensation occurs.

* *Gerlands Beiträge*, 15, 170–172. † *Loc. cit.* 175–176.

‡ This statement was qualitatively anticipated by G. R. Putnam, in a discussion of gravity in the United States (*U.S. Coast and Geodetic Survey, Report, 1894*, Appendix I). The attraction of the deep-seated compensation was replaced in this reduction by that of the average surface inequality within 160 km., with its sign changed.

11·62. The hypothesis of purely local compensation would mean, mechanically, that the addition of a load on top depressed the cylinder of matter vertically below it down to the asthenosphere, without affecting the matter outside this cylinder at all. This amounts to regarding the crust as made of vertical floating piles, each able to slide freely up and down without interference from the others. It is mechanically possible, but extremely unlikely. The probable conditions are (1) that the whole crust is coherent, and bends like a beam or an elastic plate when a load is added, or (2) that the region where the matter is added is bounded at one or all sides by fault or thrust planes, the block between them sinking or rising as one piece. In either case compensation will be regional. In the first its distribution could be found by the method of Chapter X, or by the simpler approximate methods used in the theory of the bending of beams. It is found in the latter way* that if a floating plate, of infinite length parallel to the axis of x, is bent in two dimensions by a load σg per unit area over the strip between $x = \pm\, l$, the downward displacement at any point is

$$y = \frac{\sigma}{\rho}\left[1 - \tfrac{1}{2}\, e^{-\alpha(l-x)} \cos \alpha\,(l-x) - \tfrac{1}{2}\, e^{-\alpha(l+x)} \cos \alpha\,(l+x)\right] \qquad -l < x < l \qquad \text{.........(1)},$$

$$y = \frac{\sigma}{2\rho}\left[e^{-\alpha(x-l)} \cos \alpha\,(x-l) - e^{-\alpha(x+l)} \cos \alpha\,(x+l)\right] \qquad x > l \qquad \text{.........(2)},$$

where ρ is the density of the fluid below, and

$$\alpha^4 = \frac{3g\rho}{Ed^3} \qquad \text{.........(3)},$$

where E is Young's modulus for the plate, equal to about $\frac{5}{2}\mu$ for crustal rocks, and d is the depth of the plate. This solution may be used as an approximate expression of the deformation in the early stages of the filling up of an estuary with sediments or the excavation of a river valley. The depression would be simply σ/ρ below the plate, and zero outside it, if compensation were local. Actually these relations hold approximately at distances from the margins greater than $\pm\,\frac{1}{2}\pi/\alpha$, but within a belt of width π/α centred on the margins the depression varies continuously from σ/ρ to zero.

But the bending moment at any point is $\frac{1}{12}\,Ed^3\,\dfrac{d^2y}{dx^2}$ per unit breadth, and d^2y/dx^2 is greatest when

$$\alpha\,(x-l) = \pm\,\tfrac{1}{4}\pi \qquad \text{.........(4)},$$

apart from the slight effect of the terms in $x + l$. The greatest tensile stress is $\frac{1}{2}Ed\,\dfrac{d^2y}{dx^2}$, which at its greatest is

$$0{\cdot}16\,Ed\alpha^2\,\frac{\sigma}{\rho} = 0{\cdot}28\,\sigma\left(\frac{Eg}{\rho d}\right)^{\frac{1}{2}} \qquad \text{.........(5)}.$$

* *Gerlands Beiträge*, **15**, 183–187.

Now we must take ρ in the lower layer equal to 3·4; $g = 980$; E for granite about 7×10^{11} dynes/cm.[2] The strength of granite is about 8×10^{8} dynes/cm. With d equal to 30 and 50 km. respectively this makes the maximum possible values of σ equal to 4×10^5 and 5×10^5 gm./cm.^{2} respectively, so that the depth of material added or removed cannot exceed about 2 km. Allowing for the fact that it depresses the crust by about two-thirds its thickness, supposing the added matter to have density 2·4 and the matter squeezed out below 3·4, we see that this solution will fail whenever the height of the surface inequality exceeds about 700 metres. The most uncertain factor in this estimate is d, the thickness of the lithosphere. But it cannot exceed the depth of compensation found for local compensation, and when d is less than this the permissible loading is less. The critical strength assumed is the crushing strength of granite at the surface; the strength may be greater lower down, but the greatest stress in a bent beam *is* at the surface, and the actual strength there may be lower since sedimentary rocks are involved. The theory of the bent beam therefore fails for the larger inequalities of the crust.

It should be stated that a mass 700 metres in height, of horizontal extent small compared with d, will produce a maximum stress in the lithosphere of only about 3×10^8 dynes/cm.^{2}, which could be supported easily. Our difficulty does not concern narrow valleys or ridges, but only regions where the mean elevation departs from normal over horizontal distances so great as to cause great stresses throughout the lithosphere. For large mountain systems this is true. It appears therefore that for them we must adopt the second alternative, that they are bounded by faults, thrusts, or other regions of yield, formed during their development or soon after. The compensation is then distributed over the region between the extreme planes of failure. Incidentally the result of fracture due to overloading or unloading would be faults with throws amounting to some hundreds of metres, in general agreement with those of observed faults.

11·63. On the latter view we may suppose the compensation of a surface inequality to fall off linearly with distance from it, being zero at a horizontal distance l, where the outermost fracture reaches the layer where the compensation is. But gravity due to a long narrow strip at depth c below a flat surface is equivalent to that produced by a distribution of density over this surface proportional to $c^2/(c^2 + x^2)$. This reaches a quarter of its maximum value at a distance $c\sqrt{3}$, which may be taken roughly as a measure of l.

On our theory the compensation would be concentrated at the base of the lithosphere if this was homogeneous; but if the density changes suddenly within it the compensation will appear to be at the levels where the changes occur. In our case it will be half at the bottom of the granitic layer, say 12 km. down, and half at the bottom of the intermediate layer,

36 km. down, and we may regard it as equivalent to compensation at a depth of 24 km. Now Heiskanen's depth for local compensation is 41 km. for the Alps and 77 km. for the Caucasus, that is, 17 km. and 53 km. respectively below the layer where the average compensation is. Then the compensation may be considered as spread to distances of 30 and 90 km. respectively. For the Alps this would imply thrust planes at an angle of the order of 40° to the horizontal, for the Caucasus about 15°. But obviously considerable latitude in interpretation remains. The Caucasus seem to be abnormal; other mountain ranges give results more resembling those for the Alps.

11·64. The answer to our third question, whether the outstanding residuals on any isostatic reduction are real or due to errors of observation, is in favour of the first alternative. The probable error of a determination of gravity in any of the better surveys is about 0·003 cm./sec.^{2}; the mean residual after the depth of compensation has been chosen to fit the facts best is still about 0·017 cm./sec.^{2} Of the various hypotheses that might account for the residuals, the most probable seems to be real departures from exact isostasy*. The mean residual would correspond to an extra thickness of 170 metres of uncompensated matter of density 2·5. The corresponding stress on the asthenosphere would be 4×10^7 dynes/cm.^{2} But some individual residuals may be three times as great, and the strength of the asthenosphere may be as much as 10^8 dynes/cm.^{2} The residuals are not distributed at random geographically, but keep the same sign over distances of the order of 200 km., as may be seen from the maps in Bowie's memoir of 1917; stresses due to them are therefore not borne by the lithosphere, but must be transmitted to the asthenosphere.

Rather curiously, an outstanding difficulty of isostasy comes from the country where its approximate truth was first recognized. Over the great Gangetic Plain gravity is systematically low, and it is high over the flanks of the Himalaya and Vindhya Mountains to the north and south. The residuals did not yield to any ordinary treatment until Burrard† found that they would be explained by a trough filled with alluvium and light rocks of density 2·4 to a depth of about 50,000 feet (15 km.). But such a depth seems very improbable on other grounds, and the anomalies to be explained are only about 0·030 cm./sec.^{2}, not too great to be attributed to real departures from isostasy, especially as the width involved is about 300 km.

It was thought by Barrell that large deltas, such as those of the Nile and Niger, represented additional loads on the crust and therefore would be uncompensated. These particular deltas have not been surveyed from this point of view. But Burrard remarks‡ that the gravity anomalies over the delta of the Ganges are slightly negative, and Bowie§ finds that the

* *Gerlands Beiträge*, **15**, 174–175.

† *Survey of India, Professional Paper*, **17**, 1918. ‡ *Loc. cit.* 6.

§ *U.S. Coast and Geodetic Survey*, Special Publication, **99**, 1924.

eight stations on the Mississippi delta have a mean anomaly of − 0·007 cm./sec.²

11·65. Definite evidence on the thickness of the lithosphere is lacking. Barrell gave about 300 km., but this was little more than a guess, resting on a mixture of Airy's mechanics with the hypothesis of uniform local compensation. Fractures due to loading usually tend to run at about 45° to the greatest stress, and the horizontal extent of the outflow would be expected to be comparable with the thickness in question. It may therefore be comparable with Heiskanen's depth of compensation, say 40 to 80 km., but it may be as small as 30 km.

11·66. A mountain range rising 10,000 metres above its adjacent valleys would correspond to the case where the ν of 10·6 is $1{\cdot}2 \times 10^9$ dynes/cm.², and the stress-difference needed to support it in a uniform crust would be about 10^9 c.g.s. This is near the crushing strength of basalt, which could therefore just support the Himalayas if the stresses necessary could be distributed over an infinite depth. But a weak layer below throws extra stress on the upper part of the crust, and therefore the strength of the lithosphere at some depth, even if it is fractured, must be distinctly more than that of surface basalt. This is in accordance with experimental evidence obtained by F. D. Adams and L. V. King*, who have shown that at the pressure existing at a depth of 18 km. and a temperature of 550° C., Westerly granite acquires a strength of about 6×10^9 dynes/cm.² The strength of the rocks of the crust therefore probably increases downwards till it becomes several times that of surface rocks; at some depth, probably about the base of the intermediate layer, it begins to decrease, and may be a tenth of that of surface rocks at a depth of 50 km. A variation of this type was first inferred by Barrell in a series of papers already mentioned.

11·7. *Summary.* A classification of the mechanical properties of matter in the solid and liquid states has been given. The apparent small gravitative effect of mountains is explained on Airy's lines as due to the squeezing out of a weak but dense substratum by the weight of loads added on top. It is shown that compensation distributed uniformly down to a definite depth would be consistent with Airy's mechanics, but that with the actual structure of the crust compensation is probably concentrated at the bases of the granitic and intermediate layers. It appears that a widespread visible inequality more than about 700 metres in height would lead to fractures in the lithosphere, and that great mountain ranges must be bounded by fault planes and supported by hydrostatic pressure and not by the strength of the rest of the lithosphere. The outstanding residuals seem to point to a strength in the lower layer of the order of 10^8 dynes/cm.², say one-eighth of that of granite at the surface.

* *Journ. Geol.* **20**, 1912, 97–138.

CHAPTER XII

The Figures of the Earth and Moon

"There's mair ways o' killin' a pig than by greasin' him wi' het butter." Northumbrian proverb.

12·1. In the last chapter we considered the effect of the weak layer within the crust, suggested by thermal theory, on the compensation of surface inequalities within the continents, on the scale of the great mountain ranges. We have now to extend the same principles to the earth as a whole, considering the ellipticity of its figure expressed by the difference between its polar and equatorial radii, and the differences between continents and ocean basins. Since the departures of the earth and the moon from spherical symmetry are small, it is permissible to discuss the various types of disturbance separately, and to regard the effects of all together as the sum of the portions due to each separately.

12·11. *The Gravitational Potential of a nearly Spherical Body.* We shall evidently need to know the gravitational forces due to the disturbed body itself. Let us consider first the potential due to a homogeneous mass of almost spherical form. Its density is to be ρ, supposed uniform. Take spherical polar coordinates with regard to a point near the centre. The radius vector to a point on the surface in the direction given by the angular coordinates (θ, ϕ) can be expressed in the form

$$r = a\,(1 + Y_1 + Y_2 + \ldots + Y_n + \ldots) \qquad \ldots\ldots\ldots(1),$$

where the Y's are spherical surface harmonic functions of θ and ϕ. It will be supposed that the mass is so nearly spherical that all the Y's are small enough for their squares and products to be neglected. If the gravitational potential at a point (r, θ', ϕ') be U_0 when the point is outside the body, and U_1 when it is inside, the conditions for the continuity of the potential and its normal derivative are easily seen to be satisfied, subject to the above proviso concerning the neglect of squares and products of the Y's, if

$$U_0 = \frac{4}{3}\pi f \rho a^3 \left(\frac{1}{r} + \frac{aY_1'}{r^2} + \frac{3}{5}\frac{a^2Y_2'}{r^3} + \ldots + \frac{3}{2n+1}\frac{a^nY_n'}{r^{n+1}} + \ldots\right) \qquad \ldots\ldots\ldots(2),$$

$$U_1 = \frac{4}{3}\pi f \rho a^3 \left(\frac{3a^2 - r^2}{2a^3} + \frac{rY_1'}{a^2} + \frac{3}{5}\frac{r^2Y_2'}{a^3} + \ldots + \frac{3}{2n+1}\frac{r^nY_n'}{a^{n+1}} + \ldots\right) \quad \ldots(3),$$

where the Y''s are the same functions of θ' and ϕ' that the Y's are of θ and ϕ, and f is the constant of gravitation. U_1 and U_0 are therefore the appropriate values of the internal and external potentials.

12·12. Passing now to the case of a heterogeneous body, let the equation of the stratum whose density is ρ be

$$r = r_1 (1 + Y_1 + Y_2 + \ldots + Y_n) \qquad \ldots\ldots\ldots(1),$$

where the Y's may evidently be functions of r_1. Consider the potential due to the shell between the layers where the densities are respectively ρ' and $\rho' + d\rho'$. To the first degree in $d\rho'$, the contribution to U_0 due to this shell is the difference between the potentials due to two homogeneous bodies, both of density ρ', one filling the outer of these layers and the other the inner. Then the potential due to a homogeneous body filling either stratum can be found from 12·11 (2) and (3), and the potential due to the layer between the two strata considered is

$$\frac{4}{3}\pi f \rho' \frac{\partial}{\partial a'}\left\{\frac{a'^3}{r} + \frac{a'^4 Y_1'}{r^2} + \frac{3}{5}\frac{a'^5 Y_2'}{r^3} + \ldots + \frac{3}{2n+1}\frac{a'^{n+3} Y_n'}{r^{n+1}} + \ldots\right\} da' \ldots(2),$$

where a' is the value of r_1 corresponding to $\rho = \rho'$, and other accented letters also refer to this layer. By integrating for all layers we therefore find that the potential at external points due to the whole heterogeneous body is

$$U_0 = \frac{4}{3}\pi f \int_0^a \rho' \frac{\partial}{\partial a'}\left\{\frac{a'^3}{r} + \frac{a'^4 Y_1'}{r^2} + \frac{3}{5}\frac{a'^5 Y_2'}{r^3} + \ldots + \frac{3}{2n+1}\frac{a'^{n+3} Y_n'}{r^{n+1}} + \ldots\right\} da'$$
$$\ldots\ldots\ldots(3),$$

where a is the mean radius of the outer surface of the body. We may find the potential at an internal point similarly. If the mean radius of the stratum of equal density through the point is r_1, the matter within this contributes to the potential an amount

$$\frac{4}{3}\pi f \int_0^{r_1} \rho' \frac{\partial}{\partial a'}\left\{\frac{a'^3}{r} + \frac{a'^4 Y_1'}{r^2} + \frac{3}{5}\frac{a'^5 Y_2'}{r^3} + \ldots + \frac{3}{2n+1}\frac{a'^{n+3} Y_n'}{r^{n+1}} + \ldots\right\} da' \quad (4),$$

while the matter outside of it is found from 12·11 (3) in a similar way to give

$$\frac{4}{3}\pi f \int_{r_1}^a \rho' \frac{\partial}{\partial a'}\left\{\frac{3}{2}a'^2 + a' r Y_1' + \frac{3}{5} r^2 Y_2' + \ldots + \frac{3}{2n+1}\frac{r^n}{a'^{n-2}} Y_n' + \ldots\right\} da'$$
$$\ldots\ldots\ldots(5).$$

Thus U_1 is the sum of the two expressions (4) and (5).

12·13. The mass of the heterogeneous body is evidently given by

$$M = 4\pi \int_0^a \rho' a'^2 da' \qquad \ldots\ldots\ldots(1).$$

We shall frequently have occasion to use the function

$$S(r_1) = 3 \int_0^{r_1} \rho' a'^2 da' \qquad \ldots\ldots\ldots(2).$$

Thus $S(r_1)$ is $3/4\pi$ times the mass within the stratum of equal density whose mean radius is r_1. In particular,

$$S(a) = \frac{3M}{4\pi} \qquad \ldots\ldots\ldots(3).$$

We shall also need to use the moments of inertia of the body and the differences between them. If A, B, C denote the moments of inertia about the axes of x, y, and z respectively, we have

$$A = \iiint \rho\,(y^2 + z^2)\,d\tau \qquad \text{.........(4)},$$

with two similar expressions, each triple integral being taken through the body. By subtraction,

$$C - A = \iiint \rho\,(x^2 - z^2)\,d\tau \qquad \text{.........(5)},$$

with two similar expressions. Now let us put

$$x = r \sin\theta \cos\phi \qquad \text{.........(6)},$$

$$y = r \sin\theta \sin\phi \qquad \text{.........(7)},$$

$$z = r \cos\theta \qquad \text{.........(8)}.$$

Then

$$C = \int_0^R \int_0^\pi \int_0^{2\pi} \rho r^4 \sin^3\theta\, dr\, d\theta\, d\phi \qquad \text{.........(9)},$$

where R denotes the distance from the centre to the surface in the direction given by (θ, ϕ). Suppose first that the body is homogeneous. We can write

$$\sin^2\theta = \tfrac{2}{3} + (\tfrac{1}{3} - \cos^2\theta) \qquad \text{.........(10)},$$

and the second part of this is a spherical surface harmonic of order 2. Then

$$C = \tfrac{1}{5} \int_0^\pi \int_0^{2\pi} \rho R^5 \{\tfrac{2}{3} + (\tfrac{1}{3} - \cos^2\theta)\} \sin\theta\, d\theta\, d\phi \qquad \text{.........(11)},$$

and on substituting for R from 12·11 (1) and neglecting squares of the Y's,

$$C = \tfrac{1}{5} \int_0^\pi \int_0^{2\pi} \rho a^5 (1 + 5Y_1 + 5Y_2 + \dots + 5Y_n) \{\tfrac{2}{3} + (\tfrac{1}{3} - \cos^2\theta)\} \sin\theta\, d\theta\, d\phi$$
$$\text{.........(12)}.$$

Now if S_m and S_n be any two surface harmonics of different orders we know that

$$\int_0^\pi \int_0^{2\pi} S_m S_n \sin\theta\, d\theta\, d\phi = 0,$$

and therefore when the integrand in (12) is multiplied out and integrated, the only terms that do not vanish are

$$C = \tfrac{8}{15}\pi\rho a^5 + \rho a^5 \int_0^\pi \int_0^{2\pi} Y_2 (\tfrac{1}{3} - \cos^2\theta) \sin\theta\, d\theta\, d\phi \qquad \text{......(13)}.$$

The harmonics of orders different from 0 and 2 therefore contribute nothing to the moments of inertia.

Now Y_2 must be of the form

$$Y_2 = (\alpha x^2 + \beta y^2 + \gamma z^2 + 2fyz + 2gzx + 2hxy)/r^2 \qquad \text{......(14)}.$$

When this is substituted in (13) and integrated, the product terms contain factors

$$\int_0^{2\pi} \sin\phi\, d\phi, \quad \int_0^{2\pi} \cos\phi\, d\phi, \quad \int_0^{2\pi} \sin\phi \cos\phi\, d\phi,$$

all of which vanish. The other terms in C become

$$C = \tfrac{8}{15}\pi\rho a^5 + \tfrac{8}{45}\pi\rho a^5 (\alpha + \beta - 2\gamma) \qquad \text{.........(15)}.$$

If then we retain only the largest terms, we have

$$C = \tfrac{8}{15}\pi\rho a^5 \qquad \text{.........(16)},$$

$$C - A = \tfrac{8}{15}\pi\rho a^5 (\alpha - \gamma) \qquad \text{.........(17)},$$

with symmetrical expressions for the other moments of inertia and their differences. It follows that for a homogeneous ellipsoid, if the axes are the principal axes, the ratio $(C - A)/C$ is equal to the ellipticity of the section by the plane $y = 0$.

The two results (16) and (17) can be generalized by the method of 12·12. We thus get

$$C = \tfrac{8}{15}\pi \int_0^a \rho' da'^5 \qquad \text{.........(18)},$$

$$C - A = \tfrac{8}{15}\pi \int_0^a \rho' d\{a'^5 (\alpha - \gamma)\} \qquad \text{.........(19)}.$$

Thus the ratio $(C - A)/C$ for a heterogeneous body is a weighted mean of the ellipticities of the strata of equal density.

12·2. *Condition for Equilibrium of a Rotating Fluid.* Consider a body in a state of steady rotation about a fixed axis, which will be taken to be the axis of z. Let the angular velocity be Ω. Then the component accelerations of any element of the body are $(-\Omega^2 x, -\Omega^2 y, 0)$. If now any part of the body is under no shearing stress (in other words, if it is fluid, or if it is a solid, but has had any stress-differences removed by plastic adjustment) the equations of motion will be, as in ordinary hydrodynamics,

$$-\Omega^2 x = \frac{\partial U}{\partial x} - \frac{1}{\rho}\frac{\partial p}{\partial x} \qquad \text{.........(1)},$$

$$-\Omega^2 y = \frac{\partial U}{\partial y} - \frac{1}{\rho}\frac{\partial p}{\partial y} \qquad \text{.........(2)},$$

$$0 = \frac{\partial U}{\partial z} - \frac{1}{\rho}\frac{\partial p}{\partial z} \qquad \text{.........(3)},$$

where U is the complete gravitation potential, p the pressure, and ρ the density. If we write

$$\Psi = U + \tfrac{1}{2}\Omega^2 (x^2 + y^2) \qquad \text{.........(4)},$$

the equations of motion are together equivalent to the single total differential equation

$$dp = \rho d\Psi \qquad \text{.........(5)}.$$

It follows that p is a function of Ψ, and that ρ is either the same everywhere or another function of Ψ. In particular, if the fluid or quasi-fluid has a free surface, the pressure is uniformly zero over this surface, and therefore Ψ is constant over it.

It has been seen that the whole of the interior of the earth, at depths greater than about 40 km., is probably very weak, and must be still in an approximately hydrostatic state, while before solidification this state must have held right up to the surface. Accordingly the theory just given was applicable to the whole of the earth before solidification, and

is still applicable to most of the interior. It is certain, however, that changes occurring in the crust at depths less than 40 km. have prevented the hydrostatic state from persisting at these depths, and it is possible that such changes may have influenced the adjustment of the form of the outer layers to the earth's actual speed of rotation. The discrepancy is unlikely to be large, since the thickness of these layers is only a small fraction of the radius of the earth, and they would probably bend in such a way as to adjust themselves with some accuracy to inequalities of great horizontal extent, such as the one produced by rotation; but it may be appreciable.

On the other hand, we know that the ocean is a fluid and, apart from the disturbing influences of gravitational and meteorological tides, which are small in comparison with its ellipticity of figure, the theory must hold exactly. It is therefore desirable to make what progress we can from empirical data obtainable on the surface of the ocean, before we proceed to supplement them by means of a theory of the conditions in the interior of the earth.

The surface Ψ = constant that includes the ocean is known as the 'geoid.' The spirit level at a place sets itself parallel to it, and therefore heights above sea level as measured are usually really heights above the geoid. The 'spheroid' is the ellipsoid of revolution that most closely resembles the geoid.

12·21. We notice that in general we can write

$$\tfrac{1}{2}\Omega^2 (x^2 + y^2) = \tfrac{1}{3}\Omega^2 r^2 + \tfrac{1}{2}\Omega^2 r^2 (\tfrac{1}{3} - \cos^2\theta) \qquad \text{.........(1)},$$

where θ is the colatitude, or the difference between the true latitude and 90°. Thus the effect of rotation can be expressed by the addition to the gravitation potential of two small terms, one symmetrical about the centre of the body, and the other a solid harmonic of order 2.

12·22. *The Form of the Ocean Surface.* Let us now proceed to consider the effect of rotation on the form of the ocean surface. We have seen that it may be represented by introducing a disturbing potential involving terms of orders 0 and 2. If now we consider only terms of these orders, and use the suffix a to denote the surface value of the quantity concerned, we have from 12·12 (3)

$$U_0 = \frac{4}{3}\pi f \int_0^a \rho' \frac{\partial}{\partial a'} \left(\frac{a'^3}{r} + \frac{3}{5}\frac{a'^5}{r^3} Y_2' \right) da' \qquad \text{.........(1)},$$

and the form of the surface is given by

$$r = a\,(1 + Y_{2,a}) \qquad \text{.........(2)}.$$

Substituting from (2) in (1), and neglecting terms of the second degree in Y_2, we have

$$U_{0,a} = 4\frac{\pi f}{a} \int_0^a \rho' a'^2 da' \,(1 - Y_{2,a}) + \frac{4}{5}\frac{\pi f}{a^3} \int_0^a \rho' \frac{\partial}{\partial a'} (a'^5 Y_2')\, da' \;\text{...(3)},$$

and on substituting in Ψ, and remembering 12·13 (1), we have

$$\frac{fM}{a}(1 - Y_{2,a}) + \frac{4\pi f}{5a^3}\int_0^a \rho' \frac{\partial}{\partial a'}(a'^5 Y_2')\,da' + \tfrac{1}{3}\Omega^2 a^2 + \tfrac{1}{2}\Omega^2 a^2(\tfrac{1}{3} - \cos^2\theta)$$
$$= \text{constant} \qquad \text{.........(4).}$$

The terms fM/a and $\frac{1}{3}\Omega^2 a^2$ are constant over the surface as they stand. Now Y_2' will in general be a linear function of the five different surface harmonics of the second order. Thus (4) gives, on equating coefficients of each harmonic separately, one equation satisfied by each coefficient. At present, however, we are concerned only with the harmonic $\frac{1}{3} - \cos^2\theta$. If its coefficient in Y_2 is ϵ, we have on picking out coefficients of $\frac{1}{3} - \cos^2\theta$ in (4)

$$-\frac{fM}{a}\epsilon_a + \frac{4\pi f}{5a^3}\int_0^a \rho' \frac{\partial}{\partial a'}(a'^5\epsilon')\,da' + \tfrac{1}{2}\Omega^2 a^2 = 0 \qquad \text{.........(5).}$$

Now if we ignore all other inequalities, so that the surface is

$$r = a\{1 + \epsilon_a(\tfrac{1}{3} - \cos^2\theta)\} \qquad \text{.........(6),}$$

we see that the ratio of the equatorial and polar radii of the surface is, to the first degree in ϵ_a, equal to $1 - \epsilon_a$. Thus ϵ_a is the ellipticity of the ocean surface. Similarly ϵ is the ellipticity of any other stratum of equal density. Now using 12·13 (19), we have

$$-\frac{fM}{a}\epsilon_a + \frac{3}{2}\frac{f(C-A)}{a^3} + \tfrac{1}{2}\Omega^2 a^2 = 0 \qquad \text{.........(7).}$$

Let us put

$$\Omega^2 a^3/fM = m \qquad \text{.........(8),}$$

so that m is a small number. Then (7) becomes

$$\frac{3}{2}\frac{C-A}{Ma^2} = \epsilon_a - \tfrac{1}{2}m \qquad \text{.........(9).}$$

This important result shows that if the ellipticity of the ocean surface and the ratio of the centrifugal force at the equator to mean gravity are known, it is possible to infer the difference between the principal moments of inertia of the earth.

Referring back to (1), we see that the relevant part of U_0 is

$$U_0 = \frac{fM}{r} + \frac{3}{2}\frac{f(C-A)}{r^3}(\tfrac{1}{3} - \cos^2\theta) \qquad \text{.........(10).}$$

Now the observable value of gravity is the acceleration of a freely falling body with regard to the crust below it. The acceleration relative to the centre of the earth is $\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right)U_0$; but each particle of the earth has, in consequence of the rotation, an acceleration $(-\Omega^2 x, -\Omega^2 y, 0)$. The acceleration of a falling body with regard to the earth below it is therefore

$$\left(\frac{\partial U_0}{\partial x} + \Omega^2 x, \frac{\partial U_0}{\partial y} + \Omega^2 y, \frac{\partial U_0}{\partial z}\right) = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right)\Psi \qquad \text{.........(11),}$$

by 12·2 (4). Now Ψ is constant over the ocean surface, and therefore its

gradient has no component along the surface. Thus observable gravity is normal to the ocean surface. If its value is g,

$$g^2 = \left(\frac{\partial \Psi}{\partial x}\right)^2 + \left(\frac{\partial \Psi}{\partial y}\right)^2 + \left(\frac{\partial \Psi}{\partial z}\right)^2$$

$$= \left(\frac{\partial \Psi}{\partial r}\right)^2 + \left(\frac{\partial \Psi}{r\partial \theta}\right)^2 \quad \text{.........(12).}$$

Now $\frac{\partial \Psi}{r\partial \theta}$ is a small quantity compared with $\frac{\partial \Psi}{\partial r}$, and therefore we may neglect its square and put

$$g = -\frac{\partial \Psi}{\partial r} \quad \text{.........(13),}$$

the negative sign indicating that g is considered positive when it acts in the direction of decreasing r. Now from (9) and (10), and 12·21 (1),

$$\Psi = fM\left[\frac{1}{r} + \frac{a^2}{r^3}(\epsilon_a - \tfrac{1}{2}m)(\tfrac{1}{3} - \cos^2\theta) + \frac{mr^2}{a^3}\{\tfrac{1}{3} + \tfrac{1}{2}(\tfrac{1}{3} - \cos^2\theta)\}\right] \text{...(14),}$$

and therefore

$$g = fM\left[\frac{1}{r^2} + \frac{3a^2}{r^4}(\epsilon_a - \tfrac{1}{2}m)(\tfrac{1}{3} - \cos^2\theta) - \frac{2mr}{a^3}\{\tfrac{1}{3} + \tfrac{1}{2}(\tfrac{1}{3} - \cos^2\theta)\}\right] \text{...(15),}$$

and on substituting for r from (6) and neglecting squares of ϵ_a and m,

$$g = \frac{fM}{a^2}\{(1 - \tfrac{2}{3}m) - (\tfrac{5}{2}m - \epsilon_a)(\tfrac{1}{3} - \cos^2\theta)\}$$

$$= \frac{fM}{a^2}\{1 + \alpha + (\tfrac{5}{2}m - \epsilon_a)\cos^2\theta\} \quad \text{.........(16),}$$

where α is a small quantity independent of θ. If now G denote the intensity of gravity on the equator, where $\theta = \frac{1}{2}\pi$, we evidently must have

$$G = \frac{fM}{a^2}(1 + \alpha) \quad \text{.........(17),}$$

and on dividing (16) by (17), and again rejecting squares of small quantities,

$$g = G\{1 + (\tfrac{5}{2}m - \epsilon_a)\cos^2\theta\} \quad \text{.........(18).}$$

Thus the increase of gravity in passing from the equator to some other latitude is proportional to the square of the sine of the latitude reached. The formula (18) is due to Clairaut*. It will be seen that if we know m, and can find by experiment the variation of gravity over a wide range of latitude, it will enable us to find ϵ independently of trigonometric determinations.

From a comparison of gravity determinations in many widely separated regions Bowie† has found that

$$\tfrac{5}{2}m - \epsilon_a = 0{\cdot}005294 \quad \text{.........(19).}$$

* *Théorie de la Figure de la Terre*, 1743. It has also been proved by Stokes, *Scientific Papers*, **2**, 104–130, that the formula holds without the assumption that the strata of equal density are nearly spherical, provided the ocean surface is itself an ellipsoid of revolution of small ellipticity.

† *U.S. Coast and Geodetic Survey, Spec. Publ.* **40**, 1917, 134.

Also $$1/m = 288{\cdot}4 \qquad \text{.........(20).}$$

Thus $$\epsilon_a = 0{\cdot}003373 = 1/296{\cdot}4 \qquad \text{.........(21).}$$

Bowie gets $$\epsilon_a = 1/297{\cdot}4 \qquad \text{.........(22),}$$

apparently by allowing for the second degree terms, which have here been neglected. For the sake of internal consistency, however, the value (21) will be used in the following calculation. Now (9) gives

$$\frac{3}{2}\frac{C-A}{Ma^2} = 0{\cdot}001640 \qquad \text{.........(23).}$$

The ratio $(C - A)/C$ can be found from the period of the precession of the equinoxes, and is known with a smaller probable error than the surface ellipticity. We have

$$\frac{C-A}{C} = \frac{1}{305{\cdot}6} = 0{\cdot}003272 \qquad \text{.........(24).}$$

On dividing (23) by (24) we have

$$\frac{C}{Ma^2} = 0{\cdot}3341 \qquad \text{.........(25).}$$

If the earth were exactly homogeneous, this ratio would be 0·4000. Thus we have a definite indication that the earth is denser near the centre than outside.

12·3. *The Condition for Absence of Shearing Stress Internally.* The above results depend only on the hypothesis that there is no shearing stress in the ocean. If now we introduce the further hypothesis, which we have seen in 12·2 is unlikely to be seriously in error, that, so far as inequalities of large extent are concerned, the earth is in a hydrostatic state throughout, we see that Ψ must be constant over all the surfaces of equal density within the earth. Hence if in it we substitute for r its value

$$r_1(1 + Y_1 + Y_2 + \dots + Y_n + \dots) \qquad \text{.........(1)}$$

from 12·12 (1), Ψ must reduce to a function of r_1 alone. If this is done, and we neglect squares and products of the Y's, the terms of order n give

$$-\frac{Y_n}{r_1}\int_0^{r_1} \rho' a'^2 da' + \frac{1}{(2n+1)\,r_1^{n+1}}\int_0^{r_1} \rho' \frac{\partial}{\partial a'}(a'^{n+3}Y_n')\,da' + \frac{r_1^n}{2n+1}\int_{r_1}^{a} \rho' \frac{\partial}{\partial a'}\left(\frac{Y_n'}{a'^{n-2}}\right) da' = 0 \qquad \text{.........(2),}$$

except for $n = 0$, when the condition is satisfied automatically, and for $n = 2$, when the right side has to be replaced by

$$-\frac{1}{8\pi f}\Omega^2 r_1^2\left(\tfrac{1}{3} - \cos^2\theta\right).$$

We have thus a separate linear integral equation for each Y except Y_0.

Let us first consider the harmonic corresponding to the polar flattening, so that Y_2 is equal to $\epsilon\left(\tfrac{1}{3} - \cos^2\theta\right)$, where the ellipticity ϵ of the stratum is a small quantity and a function of r_1 alone. Then the surfaces of equal density are spheroids of revolution about the axis of rotation.

On substituting this value of Y_2 into the equation for Y_2, we find

$$-\frac{\epsilon}{r_1}\int_0^{r_1}\rho' a'^2 da' + \frac{1}{5r_1^3}\int_0^{r_1}\rho'\frac{d}{da'}(a'^5\epsilon')\,da' + \frac{r_1^2}{5}\int_{r_1}^{a}\rho'\frac{d\epsilon'}{da'}da' = -\frac{1}{8\pi f}\Omega^2 r_1^2 \quad \text{.........(3).}$$

Let us multiply by r_1^3, and then differentiate with regard to r_1. We find on simplifying

$$-\left(r_1^2\frac{d\epsilon}{dr_1} + 2\epsilon r_1\right)\int_0^{r_1}\rho' a'^2 da' + r_1^4\int_{r_1}^{a}\rho'\frac{d\epsilon'}{da'}da' = -\frac{5}{8\pi f}\Omega^2 r_1^4,$$

or, dividing by r_1^4,

$$-\left(\frac{1}{r_1^2}\frac{d\epsilon}{dr_1} + \frac{2\epsilon}{r_1^3}\right)\int_0^{r_1}\rho' a'^2 da' + \int_{r_1}^{a}\rho'\frac{d\epsilon'}{da'}da' = -\frac{5}{8\pi f}\Omega^2 \quad \text{......(4).}$$

Differentiating again with regard to r_1, we have

$$\left(\frac{1}{r_1^2}\frac{d^2\epsilon}{dr_1^2} - \frac{6\epsilon}{r_1^4}\right)\int_0^{r_1}\rho' a'^2 da' + 2\left(\frac{d\epsilon}{dr_1} + \frac{\epsilon}{r_1}\right)\rho = 0 \quad \text{.........(5).}$$

We now introduce the function $S(r_1)$, given by 12·13 (2),

$$S(r_1) = 3\int_0^{r_1}\rho' a'^2 da' \quad \text{.........(6).}$$

We may at this stage replace r_1 by r without confusion. Then (5) can be written in the forms

$$\frac{d^2}{dr^2}\{S(r)\,\epsilon\} = \frac{6}{r^2}S(r)\,\epsilon + 3r^2\epsilon\frac{d\rho}{dr} \quad \text{.........(7),}$$

and

$$\frac{d^2\epsilon}{dr^2} + \frac{6\rho r^2}{S(r)}\frac{d\epsilon}{dr} - \left(1 - \frac{\rho r^3}{S(r)}\right)\frac{6\epsilon}{r^2} = 0 \quad \text{.........(8).}$$

12·31. We are now in a position to prove that the ellipticities of the strata of equal density increase steadily from the centre to the surface. The density must increase steadily inwards, partly because the heaviest materials would sink to the centre in the fluid earth, and partly because the nearer the centre the greater the pressure, and therefore the more the material is compressed. Thus $d\rho/dr$ is essentially negative. Now

$$S(r) = \int_0^r \rho' da'^3 = \rho r^3 - \int_0^r a'^3\frac{d\rho'}{da'}da' \quad \text{.........(1),}$$

and the second term of this is always negative. Hence $S(r)$ is always greater than ρr^3, and $1 - \rho r^3/S(r)$ is always positive. When r is small, the latter quantity tends to zero at least as fast as r. Let the first term in it that does not vanish be Hr^k. Then k is at least unity, and H is positive. Now let the expansion of ϵ in powers of r start with r^p. On substituting in 12·3 (8) and equating coefficients of r^{p-2} we see that

$$p(p+5) = 0 \quad \text{.........(2),}$$

whence $p = 0$ or -5. The negative root is impossible, since it would make ϵ infinite at the centre. Therefore the only admissible root of (2) is zero, and ϵ is finite at the centre, equal to A, say. If the next term in ϵ that

does not vanish is Br^s, we may substitute $A + Br^s + \dots$ for ϵ in 12·3 (8). The terms of lowest degree are $Bs\,(s+5)\,r^{s-2} - 6Hr^{k-2}A$. Since A and H are not zero, the term in B must cancel the term in AH, and therefore $s = k$; and since H is positive, B has the same sign as A. Now when r is small $\dfrac{d\epsilon}{dr}$ behaves like sBr^{s-1}, and since s is positive this must have the same sign as B, and therefore as A. Thus $\dfrac{d\epsilon}{dr}$ must have the same sign as ϵ when r is small.

Now when $d\epsilon/dr$ and ϵ have the same sign, it follows that ϵ increases numerically with r. Thus the ellipticity increases from the centre. It could only cease to increase if $d\epsilon/dr$ became zero. But if $d\epsilon/dr$ is zero, we see from 12·3 (8) that $d^2\epsilon/dr^2$ must have the same sign as ϵ. Thus for slightly greater values of r, $d\epsilon/dr$ will again have the same sign as ϵ, and ϵ will again proceed to increase numerically. Thus ϵ must increase with r right up to the surface.

12·32. It is not, however, enough that ϵ should satisfy the differential equation 12·3 (8). It must also satisfy the first and second integrals 12·3 (3) and (4) whence this equation was derived, otherwise it will not be a solution of the problem. If in 12·3 (3) we put $r_1 = a$, and use 12·13 (1) and 12·22 (8), we get

$$\int_0^a \rho' \frac{\partial}{\partial a'}(a'^5\epsilon')\,da' = \frac{5a^2M}{4\pi}(\epsilon_a - \tfrac{1}{2}m) \qquad \dots\dots\dots(1).$$

This is equivalent to 12·22 (9). The left side of (1) would be increased numerically if ϵ were everywhere the same and equal to its surface value ϵ_a. Also $\int_0^a \rho' \frac{d}{da'} a'^5 da'$ would be increased if matter was removed from the interior to the exterior to make the density uniform. Hence on both grounds

$$\left|\int_0^a \rho' \frac{d}{da'}(a'^5\epsilon')\,da'\right| < |\epsilon_a|\,\bar{\rho}a^5 \qquad \dots\dots\dots(2),$$

where $\bar{\rho}$ is the mean density. But

$$\bar{\rho} = \frac{3M}{4\pi a^3} \qquad \dots\dots\dots(3),$$

and therefore the left side of (2) is numerically less than $3a^2M\epsilon_a/4\pi$. Hence (1) shows that ϵ_a has the same sign as m, and is therefore positive. But ϵ never changes sign, and therefore the ellipticities of all strata below the surface are positive. In other words, the value of r for any stratum is greatest when θ is equal to $\frac{1}{2}\pi$; that is, all the strata of equal density are oblate.

12·33. We can employ similar methods to show that the expression for r can contain no spherical harmonics except $\frac{1}{3} - \cos^2\theta$. For by a

process similar to that used in 12·3 we can show that Y_n must satisfy the differential equation

$$\frac{d^2 Y_n}{dr^2} - n(n+1)\frac{Y_n}{r^2} + \frac{6\rho}{S(r)}\left(r^2 \frac{dY_n}{dr} + rY_n\right) = 0.$$

If when r is small Y_n behaves like r^p, we readily see that

$$p(p+5) - n^2 - n + 6 = 0.$$

In the first place, if $n = 1$, the values of p that satisfy this equation are -1 and -4. The former corresponds to the exact solution $Y_n = 1/r$, and gives only a displacement of the earth as a whole as a rigid body; the second is impossible. Therefore the only possible first harmonic displacement is irrelevant.

If n is equal to or greater than 2, there is one zero or positive value of p, and this, by an argument exactly like that of 12·31, makes Y_n increase steadily numerically with r. Then as in 12·32 we can show that when r_1 is put equal to a, the first term in 12·3 (1) is $\bar{\rho}a^2 Y_n$, while the third is zero and the second is numerically less than $\frac{1}{2n+1}\bar{\rho}a^2 Y_n$. Thus 12·3 (1) cannot be satisfied unless Y_n is zero. It follows that the harmonic in $\frac{1}{3} - \cos^2\theta$ is the only one present in the figure of the earth on the theory of hydrostatic equilibrium.

12·34. We can now prove that ϵ/r^3, which is evidently always positive, decreases steadily as r increases. In 12·3 (8) let us put

$$\epsilon = \lambda r^3 \qquad \text{.........(1)}.$$

Then we find
$$\frac{d^2\lambda}{dr^2} + 6\left(\frac{\rho r^2}{S(r)} + \frac{1}{r}\right)\frac{d\lambda}{dr} + \frac{24\rho r}{S(r)}\lambda = 0 \qquad \text{.........(2)}.$$

Now ϵ is finite when r is zero, and therefore λ behaves like r^{-3} when r is small. It therefore diminishes as r increases. It could cease to diminish only if $d\lambda/dr$ became zero; but then (2) shows that $d^2\lambda/dr^2$ would be negative, and therefore for slightly greater values of r, $d\lambda/dr$ would again be negative, and λ would continue to diminish. Thus ϵ/r^3 decreases steadily as r increases. The theorems of 12·31 and 12·34 are due to Clairaut.

12·35. If, again, we put r equal to a in 12·3 (4), we get

$$\left\{\frac{1}{a^2}\left(\frac{d\epsilon}{dr}\right)_a + \frac{2\epsilon_a}{a^3}\right\}\int_0^a \rho' a'^2 da' = \frac{5}{8\pi f}\Omega^2 \qquad \text{.........(1)},$$

where the suffix a indicates that the corresponding quantity is to be evaluated at the surface. Using now 12·13 (3) and 12·22 (8) we have

$$a\left(\frac{d\epsilon}{dr}\right)_a + 2\epsilon_a = \tfrac{5}{2}m \qquad \text{.........(2)}.$$

12·36. *Radau's Transformation.* Let us now introduce a new dependent variable η, defined by

$$\eta = \frac{d \log \epsilon}{d \log r} = \frac{r}{\epsilon} \frac{d\epsilon}{dr} \quad \text{.........(1).}$$

Then
$$\frac{d\epsilon}{dr} = \frac{\eta\epsilon}{r}; \quad \frac{d^2\epsilon}{dr^2} = \left(\frac{1}{r}\frac{d\eta}{dr} + \frac{\eta^2}{r^2} - \frac{\eta}{r^2}\right)\epsilon \quad \text{.........(2).}$$

Substituting in 12·3 (8),

$$\frac{rd\eta}{dr} + \eta^2 - \eta + \frac{6\rho r^3}{S(r)}\eta - 6\left(1 - \frac{\rho r^3}{S(r)}\right) = 0 \quad \text{.........(3).}$$

Let us now introduce the auxiliary function ρ_0, given by

$$\rho_0 = \frac{1}{r^3}\int_0^r \rho \, dr^3 \quad \text{.........(4),}$$

so that ρ_0 is the mean density of all the matter within distance r of the centre. Then (4) can be transformed to

$$\rho r^2 = \frac{1}{3}\frac{d}{dr}(\rho_0 r^3) \quad \text{.........(5),}$$

whence
$$\frac{\rho r^3}{S(r)} = 1 + \frac{1}{3}\frac{r}{\rho_0}\frac{d\rho_0}{dr} \quad \text{.........(6),}$$

and (3) becomes
$$r\frac{d\eta}{dr} + \eta^2 + 5\eta + 2\frac{r}{\rho_0}\frac{d\rho_0}{dr}(1+\eta) = 0 \quad \text{.........(7).}$$

Now we have the identity, by logarithmic differentiation,

$$\frac{\frac{d}{dr}\{\rho_0 r^5 \sqrt{(1+\eta)}\}}{\rho_0 r^5 \sqrt{(1+\eta)}} = \frac{1}{\rho_0}\frac{d\rho_0}{dr} + \frac{5}{r} + \frac{1}{2(1+\eta)}\frac{d\eta}{dr} \quad \text{.........(8).}$$

This may be used to eliminate $\frac{d\eta}{dr}$ from (7). Then

$$\frac{2\sqrt{(1+\eta)}}{\rho_0 r^4}\frac{d}{dr}\{\rho_0 r^5 \sqrt{(1+\eta)}\} = 10\left(1 + \tfrac{1}{2}\eta - \tfrac{1}{10}\eta^2\right) \quad \text{.........(9),}$$

which can be written

$$\frac{d}{dr}\{\rho_0 r^5 \sqrt{(1+\eta)}\} = 5\rho_0 r^4 \frac{1 + \frac{1}{2}\eta - \frac{1}{10}\eta^2}{\sqrt{(1+\eta)}} \quad \text{.........(10).}$$

This equation is due to Radau*. Its importance rests on the remarkable properties of the function

$$\psi(\eta) = \frac{1 + \frac{1}{2}\eta - \frac{1}{10}\eta^2}{\sqrt{(1+\eta)}} \quad \text{.........(11).}$$

We have
$$\frac{1}{\psi}\frac{d\psi}{d\eta} = \frac{1}{20}\frac{\eta(1-3\eta)}{(1+\eta)(1+\frac{1}{2}\eta - \frac{1}{10}\eta^2)} \quad \text{.........(12),}$$

so that ψ has a minimum for $\eta = 0$ and a maximum for $\eta = \frac{1}{3}$; for large values it steadily diminishes as η increases. When $\eta = 0$, $\psi = 1$ exactly; when $\eta = \frac{1}{3}$, $\psi = 1·00074$. When $\eta = 0·56$, $\psi = 0·99929$, and when $\eta = 3$, $\psi = 0·8$.

* *Comptes Rendus*, **100**, 1885, 972–977. (Cf. Tisserand, *Mécanique Céleste*, **2**, 225.)

Now on referring back to the result of 12·34, we can write

$$\frac{r^3}{\epsilon}\frac{d}{dr}\left(\frac{\epsilon}{r^3}\right) > 0 \qquad \text{.........(13)},$$

or

$$\frac{3}{r} - \frac{1}{\epsilon}\frac{d\epsilon}{dr} > 0 \qquad \text{.........(14)}.$$

Hence, by the definition of η, η is essentially less than 3.

Again, 12·35 (2) can be written

$$\eta_a = \frac{5}{2}\frac{m}{\epsilon_a} - 2$$

$$= 0{\cdot}569 \qquad \text{.........(15)},$$

by the data of 12·22.

When $r = 0$, ϵ is finite, and $d\epsilon/dr$ is not infinite. Hence η is zero. It follows that in the earth, ψ is unity at the centre, rises to 1·00074 at the stratum where $\eta = \frac{1}{3}$, and sinks to 0·99929 at the surface. Except in the very improbable event* that η can make a wide excursion beyond the limits it attains at the ends of the range, it follows that ψ can never differ from unity by more than 8 parts in 10,000. Thus with an accuracy of this order

$$\frac{d}{dr}\{\rho_0 r^5 \surd(1+\eta)\} = 5\rho_0 r^4 \qquad \text{.........(16)}.$$

This result has been used by Darwin* to approximate to the moment of inertia of the earth. For

$$C = \tfrac{8}{3}\pi\int_0^a \rho r^4\,dr \qquad \text{.........(17)},$$

neglecting small quantities; and on replacing ρ by ρ_0 by means of (5)

$$C = \tfrac{8}{9}\pi\int_0^a\left(3r^4\rho_0 + r^5\frac{d\rho_0}{dr}\right)dr$$

$$= \tfrac{8}{9}\pi\left[\rho_{0a}a^5 - 2\int_0^a \rho_0 r^4\,dr\right] \qquad \text{.........(18)},$$

on integrating the second term by parts. But by integration of (16)

$$\int_0^a \rho_0 r^4\,dr = \tfrac{1}{5}\rho_{0a}a^5\surd(1+\eta_a) \qquad \text{.........(19)}.$$

Thus

$$C = \tfrac{8}{9}\pi\rho_{0a}a^5\{1 - \tfrac{2}{5}\surd(1+\eta_a)\} \qquad \text{.........(20)}.$$

But evidently ρ_{0a} is the mean density, so that

$$M = \tfrac{4}{3}\pi\rho_{0a}a^3 \qquad \text{.........(21)}.$$

Thus by division

$$\frac{C}{Ma^2} = \tfrac{2}{3}\{1 - \tfrac{2}{5}\surd(1+\eta_a)\} \qquad \text{.........(22)}.$$

* It was actually proved by Callandreau that η increases steadily from $r = 0$ to $r = a$ if $d^2\rho/dr^2$ is everywhere positive. This condition is satisfied by Laplace's and Roche's hypotheses (see below), by not by Wiechert's. The latter, in its original form, is actually found to make η monotonic; but if α of 12·45 is less than $1/\surd 2$, $d\eta/dr$ changes sign. (Jeffreys, *M.N.R.A.S. Geoph. Suppl.* **1**, 1924, 121–124.) The value now indicated by seismology is less than this; but allowance for increase in density with depth within the layers will tend to restore the validity of the approximation.

† *Scientific Papers*, **3**, 78–118; or *M.N.R.A.S.* **60**, 1900, 82–124.

Combining this with 12·22 (9) we get

$$\frac{C-A}{C}=\frac{\epsilon_a-\frac{1}{2}m}{1-\frac{2}{5}\sqrt{(1+\eta_a)}} \qquad \text{.........(23)}.$$

Now η_a is a known function of ϵ_a and m, by (15). Thus (23) is a relation connecting $(C-A)/C$, ϵ_a, and m, and involving no hypothetical part. Given any two of these, it should therefore be possible to calculate the third. Of the three, ϵ_a is the least accurately known from other sources. Taking then as our data

$$\frac{C-A}{C}=0{\cdot}003272 \qquad \text{.........(24)},$$

$$m=0{\cdot}003467 \qquad \text{.........(25)},$$

we may solve by putting

$$\epsilon_a=m(1-\delta) \qquad \text{.........(26)},$$

giving with the aid of (15),

$$0{\cdot}9436=\frac{\frac{1}{2}-\delta}{1-\frac{2}{5}\sqrt{\frac{5}{2(1-\delta)}-1}} \qquad \text{.........(27)},$$

which can be solved for δ by expanding and neglecting δ^3. We find

$$\delta=0{\cdot}0319 \qquad \text{.........(28)},$$

whence

$$\epsilon=\frac{1}{297{\cdot}9} \qquad \text{.........(29)},$$

agreeing with 12·22 (21) within a quantity of the order of ϵ^2.

In Darwin's monumental paper quantities of the order of the squares of the ellipticity have been retained. He gets $\epsilon = 1/296{\cdot}4$, but De Sitter, in a recalculation*, corrects this to $1/296{\cdot}92$.

12·37. *Observed Values of the Ellipticity.* At this stage it will be well to collect the chief estimates of the earth's ellipticity that have hitherto been made. Clarke's famous trigonometrical determination of 1866 gave $1/294{\cdot}9$. Hayford, in 1909, from the trigonometric survey of the United States, obtained the value $1/297{\cdot}0 \pm 0{\cdot}5$. This value was inferred from only a limited region of the earth's surface, and to that extent is imperfect; on the other hand, it is superior to all other trigonometric results in that isostasy was taken into account in reducing the observations. Helmert, from measurements of the intensity of gravity, inferred that the ellipticity was $1/298{\cdot}3 \pm 0{\cdot}7$. Hayford and Bowie†, from gravity measurements in the United States, got $1/298{\cdot}4 \pm 1{\cdot}5$. Helmert‡, in 1915, revised his earlier estimate to $1/296{\cdot}7 \pm 0{\cdot}6$. Bowie§, from a comparison of gravity observations all over the earth, got $1/297{\cdot}4$. Heiskanen|| gets $1/(296{\cdot}7 \pm 0{\cdot}6)$. The agreement between theory and the best modern observational results is within the order of magnitude both of the square of the ellipticity and of the probable errors of the experimental determinations. If the difference

* *Bull. Astr. Inst. Netherlands*, **2**, 1924, 97–108.

† *U.S. Coast and Geodetic Survey Special Publication*, No. 10, 1912.

‡ *Sitzber. d. k. Preuss. Akad. d. Wiss.* **41**, 1915, 676–685. § Cf. p. 210. || Cf. p. 195.

is real, it must be attributed to the earth's not being in a hydrostatic state, and therefore gives an estimate of the extent of the departure from that state.

12·4. *Possible Distributions of Internal Density.* We have seen that a knowledge of the values of the precessional constant and m, the ratio of the centrifugal force to gravity at the equator, is enough to fix the surface ellipticity, by 12·36 (23), and that then the surface ellipticity, the precessional constant, and m are together enough to fix the moment of inertia of the earth about its axis, by 12·36 (22). The only opportunity of confronting the theory with fact is provided by the surface ellipticity, and the agreement is satisfactory. It appears, therefore, that the form of the law of density in the interior of the earth is not a matter of importance to the theory of the figure of the earth; it cannot affect the surface ellipticity by more than a few parts in 10,000, when the precessional constant and m are known, as they are. Conversely, if we have a hypothetical distribution of density within the earth that gives the correct moment of inertia, $0{\cdot}334Ma^2$, then 12·36 (22) determines η_a, and from this 12·36 (15) determines m/ϵ_a and hence ϵ_a when m is known. Finally 12·36 (23) determines the precessional constant. Every determination is unique, and therefore so long as a hypothetical law of density gives the correct moment of inertia, it is bound to give the correct ellipticity of the surface, the correct precessional constant, and the correct variation of gravity over the surface. These are all the data we have relevant to the chief second harmonic inequality in the figure of the earth.

12·41. The above theory was developed independently by Darwin and Callandreau*, second degree terms being included. It has very much reduced the interest of special hypotheses concerning the distribution of density within the earth. Several such hypotheses have been framed, but now that it is realized that any law of density that gives the moment of inertia correctly will satisfy all the data equally well, and as nearly as observation can test, there is no longer much reason for elaborate discussion of particular laws. In this section, therefore, the chief suggested laws will be merely outlined. Full accounts of those of Laplace and Roche may be found in Tisserand's *Mécanique Céleste*; Wiechert's law is of special interest, and will be treated at somewhat greater length. The first requirement in framing any of these laws was that it should make the ellipticity of a stratum of equal density expressible in finite terms as a function of its mean radius; in other words, that it should make 12·3 (8) integrable in finite terms.

12·42. Laplace's hypothesis was that

$$3r^2\frac{d\rho}{dr} = -q^2 S(r) \qquad \text{.........(1)},$$

* *Ann. Observ. Paris,* 1889, 1–84; *Bull. Astron.* **14**, 1897, 217.

where q is a constant. Then 12·3 (7) becomes

$$\frac{d^2}{dr^2}\{S(r)\,\epsilon\} + \left(q^2 - \frac{6}{r^2}\right) S(r)\,\epsilon = 0 \qquad \text{.........(2).}$$

The solution of (1) that remains finite at the centre is

$$\rho = \frac{Q}{r}\sin qr \qquad \text{.........(3),}$$

where Q is a further constant. Then

$$S(r) = \frac{Q}{q^2}(\sin qr - qr\cos qr) \qquad \text{.........(4),}$$

$$S(r)\,\epsilon \propto \left(1 - \frac{3}{q^2 r^2}\right)\sin qr + \frac{3}{qr}\cos qr \qquad \text{.........(5).}$$

The second solution of (2) makes ϵ infinite at the centre.

When q and Q are chosen so as to make the mass and the moment of inertia correct, the surface density is about 2·8 and the density at the centre about 11.

12·43. Roche's hypothesis is

$$\rho = \sigma(1 - kr^2) \qquad \text{.........(1),}$$

where σ, k are constants.

It makes the equation for the ellipticity soluble in terms of hypergeometric functions. It gives a central density of 10·10 and a surface one of 2·3.

12·44. Laplace's and Roche's distributions of density are not physical laws in any sense. They rest on no known data about the constitution and properties of the matter within the earth. Indeed our available knowledge indicates that the interior is so dense as to be probably metallic, while the outside is rocky. Such materials cannot mix freely, and therefore we should expect a fairly sharp boundary between them, with a sudden discontinuity in density. Laplace's and Roche's formulae, however, make the density distribution continuous at all levels. It is necessary to emphasize this point, for each of these formulae, especially Laplace's, is often mentioned as if it corresponded to a true physical law describing the compressibility of the interior of the earth, which is not the case; the only reason for any interest in either formula is that it makes equation 12·3 (8) integrable in finite terms. In addition both, especially Roche's, make the density near the surface much too small, as De Sitter pointed out.

12·45. A distribution of density of greater physical interest is that of Wiechert*. In this the earth is supposed composed of two strata of uniform density, the inner and denser material extending from the centre most of the way to the surface, and the lighter occupying the outer regions. There is a marked difference between the densities of the two layers.

* *Gött. Nach.* 1897, 221–243. Kelvin and Tait (*Natural Philosophy*, 2, 420) gave the equations satisfied by the ellipticity to the first order, but did not proceed to a numerical solution.

In discussing this hypothesis a slight change of notation will be convenient. Let a and a_1 be the outer radius and the radius of the denser core respectively, and put

$$a_1 = \alpha a \qquad \qquad \text{.........(1).}$$

Let the densities of the shell and the core be ρ_0 and ρ_1, and put

$$\rho_1 = \rho_0 (1 + \mu) \qquad \qquad \text{.........(2).}$$

Then the mass is given by

$$M = \tfrac{4}{3}\pi\rho_0 a^3 (1 + \mu\alpha^3) \qquad \qquad \text{.........(3),}$$

and the moment of inertia by

$$C = \tfrac{8}{15}\pi\rho_0 a^5 (1 + \mu\alpha^5) \qquad \qquad \text{.........(4).}$$

Thus our equation 12·22 (25) gives

$$\frac{2}{5}\frac{1 + \mu\alpha^5}{1 + \mu\alpha^3} = 0{\cdot}334 \qquad \qquad \text{.........(5),}$$

and, by what has been said in 12·4, if we can find values of α and μ that will satisfy this equation, all the other data will be satisfied automatically. Now (5) can be transformed into

$$\frac{1}{\mu} = 5\alpha^3 - 6\alpha^5 \quad \text{approximately ....(6).}$$

By our physical conditions μ must be positive and finite. Hence in order that there may be any solution $6\alpha^2$ must be less than 5. Thus α must be less than 0·913. The corresponding value of ρ_1, which is the least possible for a given mean density, is 1·31 times that density.

In Wiechert's actual solution, α was equal to 0·779. This, with the above data, makes μ equal to 1·63. If now we take the mean density of the earth equal to 5·53, following Boys and Braun, we find

$$\rho_0 = 3{\cdot}12, \quad \rho_1 = 8{\cdot}22.$$

Wiechert's results were $\rho_0 = 3{\cdot}200$, $\rho_1 = 8{\cdot}206$, calculated on the supposition of a mean density of 5·58. If a factor is applied to correct the mean density, his results become $\rho_0 = 3{\cdot}17$, $\rho_1 = 8{\cdot}15$. The discrepancies are probably attributable partly to differences in the numerical data used and partly to the fact that Wiechert included terms of the second degree in the ellipticity in his analysis. The total number of possible solutions is evidently infinite.

The value adopted for ρ_0, the density of the outer shell, corresponds roughly with those of rather dense rocks, and that for the core with those of many common metals, notably iron, copper, and nickel. But actually the whole of the earth except the very surface is compressed by the weight of the overlying matter; in each layer the density must increase with depth both because the heaviest materials would tend to settle downwards initially and because the deepest parts would be the most compressed. If we assume from Gutenberg's seismological results that $\alpha = 0{\cdot}545$, thus

identifying the central core that does not transmit distortional waves with the metallic core, equation (6) gives, with a mean density of 5·53,

$$\mu = 1{\cdot}821; \ \rho_0 = 4{\cdot}27; \ \rho_1 = 12{\cdot}04.$$

The densities found, considered as means for the shell and core, are greater than those found by Wiechert, but not unreasonably so. The pressure halfway to the centre is of order $1{\cdot}2 \times 10^{12}$ dynes/cm.^{2} The bulk-modulus of the lower layer of seismology is also about $1{\cdot}2 \times 10^{12}$ dynes/cm.^{2} Thus it would be expected that if the shell was compressed merely by its own weight its mean density would be over 4 gm./cm.^{3} Similarly from the estimated pressure of $3{\cdot}2 \times 10^{12}$ dynes/cm.^{2} at the centre, with the known velocity of 9 km./sec. for compressional waves, we can show that pressure would increase the density of iron from 8 to about 12.

The question has been investigated in more detail by Williamson and L. H. Adams*. They proceeded by finding the ratio of the bulk modulus to the density at all depths from the seismological data and using the method of successive approximation. The density within each layer being supposed to vary only through compression, the distribution of density could be found. But they assume the existence of a thick layer of pallasite, a mechanical mixture of iron and olivine, between depths of 1600 and 3100 km., in order to make a continuous transition from shell to core. We now know that the transition is sudden, so that their results are no longer of much interest except as confirming generally the amounts of the increase of density due to compression and providing a method for re-calculating the distribution of density.

We may look at the matter from the other end. The mean density and moment of inertia of the earth provide estimates of the mean densities in the shell and the core. The mean pressures within them and the bulk moduli can therefore be estimated, and hence so can the reduction of density if these pressures were taken off. The results indicate that the densities would be near to those of dunite and metallic iron, and confirm our impressions that the rocky shell is of one material from a depth of 30 km. down to the great discontinuity, and that the central core is of metallic iron.

12·5. So far the only departure from the spherical form that we have considered fully is the ellipticity of figure. It was shown in 12·33 that if the earth was in hydrostatic equilibrium throughout this would be the only inequality present, and the earth's solid surface would be an ellipsoid of revolution covered by an ocean of uniform depth. But other inequalities exist and are expressed by the differences between continents and oceans and between mountains and plains. Our study of the earth's thermal history leaves it probable that there is a layer of weakness below the oceans like that below the continents, though perhaps not so much so.

* *J. Wash. Acad. Sci.* **13**, 1923, 413–428.

The greater inequalities would therefore be expected to be isostatically compensated.

Suppose that an outer layer between

$$r = a\,(1 + Y_n) \qquad \text{.........(1)}$$

and an inner boundary $\quad r = (a - h)\,(1 + kY_n) \qquad$ ....(2)

has density ρ, and floats on a lower layer in hydrostatic equilibrium. The density of the latter just below the boundary is ρ_0. Then, by 12·2 (5), Y_n must make no contribution to the gravitation potential for $r < a - h$. By 12·11 the terms arising from Y_n are

$$\frac{4}{2n+1}\pi f \rho a^3 \frac{r^n Y_n'}{a^{n+1}} + \frac{4}{2n+1}\pi f(\rho_0 - \rho)(a-h)^3 \frac{r^n k Y_n'}{(a-h)^{n+1}} \qquad \text{.........(3)},$$

and the isostatic condition is

$$k = -\left(\frac{a-h}{a}\right)^{n-2} \frac{\rho}{\rho_0 - \rho} \qquad \text{.........(4)}.$$

The contribution to the external gravitational potential is

$$\frac{4}{2n+1}\pi f \rho a^3 \frac{a^n Y_n'}{r^{n+1}} + \frac{4}{2n+1}\pi f(\rho_0 - \rho)(a-h)^3 \frac{(a-h)^n k Y_n'}{r^{n+1}}$$

$$= \frac{4}{2n+1}\pi f \rho a^{n+3}\left[1 - \left(\frac{a-h}{a}\right)^{2n+1}\right]\frac{Y_n'}{r^{n+1}} \qquad \text{.........(5)}.$$

The contribution to gravity, evaluated at $r = a$, is

$$\frac{4(n+1)}{2n+1}\pi f \rho a \left[1 - \left(\frac{a-h}{a}\right)^{2n+1}\right] Y_n' = g' \qquad \text{.........(6)},$$

and the disturbance of the gravitation potential is

$$\frac{4}{2n+1}\pi f \rho a^2 \left[1 - \left(\frac{a-h}{a}\right)^{2n+1}\right] Y_n' = U' \qquad \text{.........(7)},$$

say. The elevation of the geoid is the expression (7) divided by g. It is therefore

$$\frac{3}{2n+1}\frac{\rho}{\bar{\rho}} a \left[1 - \left(\frac{a-h}{a}\right)^{2n+1}\right] Y_n' \qquad \text{.........(8)}.$$

If we neglect squares of Y_n, gravity falls by $2g/a$ per unit increase in height, and from (6) and (8) the disturbance of gravity at geoid level is

$$g' - \frac{U'}{g}\cdot\frac{2g}{a} = g' - \frac{2U'}{a}$$

$$= \frac{4(n-1)}{2n+1}\pi f \rho a \left[1 - \left(\frac{a-h}{a}\right)^{2n+1}\right] Y_n' = g'' \qquad \text{.........(9)}.$$

The effect of an uncompensated inequality of the same height would be simply

$$\frac{4(n-1)}{2n+1}\pi f \rho a Y_n' \qquad \text{.........(10)}.$$

The effect of compensation is to reduce this effect in the ratio

$$1 - \left(\frac{a-h}{a}\right)^{2n+1} \qquad \text{.........(11)}.$$

By (10) a kilometre of material of density 2·5 would make a contribution

to gravity of 0·10 cm./sec.^{2} for harmonics of large order, 0·057 cm./sec.^{2} for $n = 3$, 0·04 cm./sec.^{2} for $n = 2$, and 0·00 cm./sec.^{2} for $n = 1$.

The average height of the continents is about a kilometre and the average depth of the oceans is about three. If this difference was uncompensated it would produce a systematic difference in gravity of about 0·4 cm./sec.^{2} But h is only about 40 kilometres, and for harmonics of low order, corresponding to widespread inequalities, (11) is nearly

$$(2n + 1)(h/a);$$

for $n = 2$ it is $\frac{1}{30}$, for $n = 3$, $\frac{1}{20}$. An uncompensated inequality 4 km. in height will give an anomaly of gravity of 0·23 cm./sec.^{2} for $n = 3$, 0·16 cm./sec.^{2} for $n = 2$, and 0 for $n = 1$. The corresponding figures for compensated inequalities will be 0·012 for $n = 3$, 0·005 for $n = 2$, and 0 for $n = 1$. Anomalies of gravity of continental extent larger than the latter amounts must be attributed to real departures from isostasy.

12·51. Search has been made for inequalities of this type in three cases. The present writer, considering the elevations of the solid surface above and below sea level, with a correction to allow for the weight of the water, estimated that the coefficient of $P_2(\cos\theta)$ in this difference was -669 ± 30 feet, or 203 ± 9 metres*. If this is uncompensated it would produce a disturbance of gravity with coefficient 0·08 cm./sec.^{2}; if compensated, 0·003 cm./sec.^{2} The corresponding differences between polar and equatorial gravity would be 0·12 and 0·005 cm./sec.^{2} Heiskanen's result for the variation of gravity with latitude, however, has a probable error of 0·012 cm./sec.^{2}, so that the second of these at least could not be detected. It seems improbable from other considerations that this inequality is mainly uncompensated.

Heiskanen finds a systematic variation of gravity with longitude. The term in gravity is $(0{\cdot}027 \pm 0{\cdot}006)\sin^2\theta \cos 2(\phi - 18^\circ \pm 5^\circ)$ where θ is the colatitude and ϕ the east longitude. This is far too large to be explained by any compensated inequalities and must apparently be attributed to a real imperfection of isostasy. It is so situated as to have maxima near the longitudes of Central Africa and Hawaii, and minima near those of Sumatra and the Andes. Its amplitude corresponds to a distribution of uncompensated matter of density 2·5 and maximum thickness 0·7 km., of second harmonic type†. Such a load on a homogeneous elastic earth

* *Memoirs R.A.S.* **60**, 1915, 187–217.

† Heiskanen gives 690 metres for the difference between the two equatorial *axes*. My 0·7 km. is the maximum departure of the semiaxis from the mean, and therefore apparently corresponds to an ellipticity of the equator four times that found by him. The reason is that his difference refers to the geoid, and is calculable from the gravity observations without further assumption; mine refers to the solid surface, which is what interests us directly, and naturally the ellipticity of the solid surface is greater than the ellipticity it induces in the geoid. Heiskanen in a later paper has included Vening Meinesz's oceanic observations in his reductions. They make an appreciable change but do not alter the general features of the inequality. Cf. *Gerlands Beitr.* **19**, 1928, 356–377.

would give a set of stress-differences in the interior* reaching about 6×10^7 dynes/cm.^{2} at the centre and 2×10^7 dynes/cm.^{2} near the surface at the equator. In the actual earth the stresses near the surface will be greater.

Heiskanen's data, however, refer only to continental stations, and the introduction of his longitude term only reduces the standard residual by one part in 16. I think, with Lenox-Conyngham†, that the representation of the outstanding gravity anomalies by a term of this type is premature. If it is real it would require a harmonic analysis over the whole earth to evaluate it; the continental data alone could probably be fitted as well or better by a set of third or fourth harmonics. The real interest of the term is that it implies the existence of systematic anomalies in gravity of continental extent.

Duffield and Vening Meinesz have made determinations of gravity over the oceans. Their results of course refer to geoid level. Duffield tried several statical methods, such as measuring the extension of a spring when a standard body is hung from it, and comparing the atmospheric pressures as recorded by mercury and aneroid barometers‡. Meinesz used a pendulum method. The disturbing effects of waves were reduced partly by working in a submarine, which was submerged to such a depth that wave motion was much reduced, and partly by using a combination of pendulums so arranged that the motion of the vessel affected all equally and left the difference between their displacements undisturbed. The two methods have not been compared over the same route. Meinesz, travelling from Holland to Batavia by way of Panama, found systematically positive anomalies averaging $+ 0{\cdot}025$ cm./sec.^{2} in the Atlantic and $+ 0{\cdot}038$ cm./sec.^{2} in the Pacific§. Duffield travelled between England and Australia by way of the Cape. The anomalies found were systematically negative over the oceans by amounts averaging about $- 0{\cdot}01$ cm./sec.^{2}, to judge by the diagrams published, but Duffield expresses doubts as to whether the methods have yet yielded as good results as they are capable of.

The oceans traversed by both investigators reached depths of the order of 5000 metres over a large part of the routes. Such an inequality of level, if uncompensated, would give widespread gravity anomalies of the order of $- 0{\cdot}5$ cm./sec.^{2} Since these are not observed we must infer that in general the difference between oceans and continents is compensated, and therefore that the ocean bottoms are of denser materials. We have thus definite evidence of a fundamental and general difference between oceanic and continental rocks.

On the other hand marked correlations are found between the depth of the sea and the anomaly of gravity. At the Philippines, for instance, a rise of level of the sea bottom from 8000 metres to sea level corresponds

* Cf. Sir G. H. Darwin, *Scientific Papers*, 2, 479. A factor e is omitted in two of the formulae.

† *Geog. Journ.* 71, 1928, 157.

‡ *M.N.R.A.S. Geoph. Suppl.* 1, 1924, 161–204.

§ *Geog. Journ.* 71, 1928, 144–160.

to a change of the gravity anomaly (reduced to sea level) from $-0{\cdot}18$ to $+0{\cdot}25$ cm./sec.[2] The corresponding change in the Bouguer anomaly would be about 0·6 cm./sec.^{2}, on the basis of rocks of density 3 being replaced by water of density 1. The existence of a change in the anomaly at all indicates that the 'free air' hypothesis is incorrect, and some form of compensation is implied. On the other hand isostatic reduction on the basis of uniform compensation to a depth of 100 km. failed to remove systematic anomalies, and it seems that compensation of local inequalities under the oceans is at a greater depth than below the continents—as was suggested above.

Attention may be called also to the fact that disturbances of gravity at geoid level, as indicated by (10) and (9) of 12·5, both contain a factor $n-1$. A harmonic inequality of the first order in gravity at geoid level cannot exist, whether the inequalities giving rise to it are compensated or not*. A systematic excess of gravity over the Pacific as a whole, which covers half the earth's surface, would probably imply such an inequality, and would be very surprising.

12·6. We can now assume the general compensation of the oceans as a first approximation to the truth and see what it implies about the structure of the ocean floor. Taking the typical continental structure so represented by 12 km. of granite of density 2·6, and 24 km. of tachylyte of density 2·9, resting on dunite of density 3·3, we see that the upper layers could be balanced by a column of dunite of thickness 30·6 km., giving a difference of level of 5·4 km., or by one of crystalline basalt of density 3·0 and thickness 33·7 km., giving a difference of level of 2·3 km. These estimates require some increase to allow for the sedimentary layer, which is mainly confined to the continents and has an average thickness of the order of a few kilometres. Also the weight of the water must be taken into account. The average height of the continents above sea level being taken as 0·5 kilometre, our first hypothesis would put the ocean bottom at a depth of 4·9 km., and it would be further depressed by 4·9/2·3 km. by the weight of the water, giving a depth of 7·0 km. in all. For a basaltic ocean floor the extra depression is 1·5/2·0 km., giving a depth of 2·7 km. The average depth of the Pacific seems to be about 5 km., and that of the Atlantic about 4 km., so that the truth probably lies between our extremes.

But both the great oceans show considerable variations in depth. The greatest depths in them are about 8 km. in the Japan and Tonga deeps, and 9–10 km. in the Guam and Philippine deeps. Such irregularities cannot be compensated locally on the scheme just given, but the regions are very restricted, having widths of the order of only 100–200 km. at most, and it seems quite possible that they may be supported regionally.

Local compensation would require a denser rock than dunite, and

* *Proc. Roy. Soc.* A, **100**, 1921, 133.

Holmes has suggested eclogite; but I think this suggestion unnecessary, and it introduces further complications.

At the same time the oceans show large shallow regions, especially near the middle, with depths of 2–3 km. It is on these that most of the oceanic islands are. They would agree with the hypothesis of 34 km. of basalt in the upper layers, but Holmes's suggestion of a superficial layer of syenite* may be applicable here. I think the difficulty he mentions about explaining how high mountains of basaltic composition manage to exist on oceanic islands such as Hawaii disappears when we consider their small horizontal extent, the frequent proximity of a deep, and the probable regional character of the compensation.

12·7. *Figure of the Moon.* Let us now proceed to consider the second harmonic inequalities in the figure of the moon. The moon's rotation about its axis must produce an oblateness of its figure; but in addition the attraction of the earth tends to raise two tidal protuberances, one just under the earth, and the other just opposite to it. Both phenomena need to be considered in any account of the figure of the moon. It will be sufficient in this discussion to treat the moon as homogeneous.

Let axes of x, y, and z be taken at the centre of the moon and moving with it. The x axis points away from the earth, the z axis is perpendicular to the plane of the orbit, and the y axis is perpendicular to both of these. If M be the mass of the earth, m that of the moon, and n the mean angular velocity of the moon about the earth, the chief part of the potential due to the earth's attraction is

$$U = fM/R \qquad \text{.........(1)},$$

where R is the distance from the point considered to the centre of the earth. If c be the moon's mean distance from the earth, we have

$$R^2 = (c+x)^2 + y^2 + z^2 \qquad \text{.........(2)},$$

and therefore
$$U = \frac{fM}{c} - \frac{fMx}{c^2} + \frac{fM}{2c^3}(2x^2 - y^2 - z^2) \qquad \text{.........(3)},$$

to the second order in the radius of the moon. Now we are supposing the motion to be one of steady revolution, the moon always keeping the same face towards the earth, so that the motion of each part of the moon is one of revolution with angular velocity n about an axis perpendicular to the plane of the orbit through the centre of gravity of the earth and moon together. The distance of the latter point from the centre of the moon is $Mc/(M+m)$, or $c/(1+\mu)$, where

$$\mu = m/M \qquad \text{.........(4)}.$$

The perpendicular distance of a point from the axis of the general rotation is therefore $\left\{\left(\frac{c}{1+\mu} + x\right)^2 + y^2\right\}^{\frac{1}{2}}$. Hence by 12·2 the effect of the steady

* *Geol. Mag.* 63. 1926. 313.

rotation and revolution in distorting the moon is equivalent to that of a potential

$$\tfrac{1}{2}n^2\left\{\left(\frac{c}{1+\mu}+x\right)^2+y^2\right\} \qquad \text{.........(5).}$$

Combining this with (3), we find that the total effective disturbing potential is

$$\frac{fM}{c}-\frac{fM}{c^2}x+\frac{fM}{2c^3}(2x^2-y^2-z^2)+\tfrac{1}{2}n^2\frac{c^2}{(1+\mu)^2}+\frac{n^2c}{1+\mu}x+\tfrac{1}{2}n^2(x^2+y^2) \qquad \text{.........(6),}$$

and then, since $n^2c^3=f(M+m)$

the two terms in x cancel, and we have altogether a disturbing potential

$$\frac{fM}{2c^3}(2x^2-y^2-z^2)+\tfrac{1}{2}n^2(x^2+y^2)$$

due to the earth and rotation together. We require to know what effect this will have on the figure of the moon.

In the first place, it is clear that the superposition of a small disturbing potential, symmetrical about the centre of the moon, will not affect its ellipticities to the first order of small quantities. Let us include, therefore, a potential $\lambda(x^2+y^2+z^2)$, λ being afterwards determined so as to make the whole potential a solid harmonic. The requisite condition for this is that

$$n^2+3\lambda=0.$$

The disturbing potential is therefore equivalent to

$$\frac{1}{2}\frac{n^2}{1+\mu}(2x^2-y^2-z^2)+\tfrac{1}{6}n^2(x^2+y^2-2z^2).$$

If we now omit μ, which is about 1/82, this reduces to

$$\tfrac{1}{6}n^2(7x^2-2y^2-5z^2) \qquad \text{.........(7),}$$

which we may write simply Kr^2S_2.

Now suppose the surface of the moon to become deformed until its equation is

$$r=a(1+\epsilon S_2) \qquad \text{.........(8).}$$

We see from 12·11 (2) that its external potential becomes

$$fm\left(\frac{1}{r}+\frac{3}{5}\frac{a^2\epsilon S_2}{r^3}\right) \qquad \text{.........(9).}$$

The condition that the moon should be in hydrostatic equilibrium is that the sum of its own external potential and the disturbing potential shall be constant over its surface. Hence $fm\left(\frac{1}{r}+\frac{3}{5}\frac{a^2\epsilon S_2}{r^3}\right)+Kr^2S_2$ must reduce to a constant when (8) is satisfied. This gives

$$\epsilon=\frac{5}{2}K\frac{a^3}{fm} \qquad \text{.........(10),}$$

and the equation of the moon's surface is

$$r = a\left\{1 + \frac{5}{12} n^2 \frac{a}{fm}(7x^2 - 2y^2 - 5z^2)\right\}$$

$$= a\left\{1 + \frac{5}{12}\frac{M}{m}\frac{a^3}{c^3}\frac{7x^2 - 2y^2 - 5z^2}{a^2}\right\} \qquad \text{.........(11).}$$

Thus the semiaxes of x, y, z on the moon are respectively

$$a\left(1 + \frac{35}{12}\frac{M}{m}\frac{a^3}{c^3}\right),\quad a\left(1 - \frac{10}{12}\frac{M}{m}\frac{a^3}{c^3}\right),\quad a\left(1 - \frac{25}{12}\frac{M}{m}\frac{a^3}{c^3}\right) \quad \text{...(12).}$$

If then A, B, C be the principal moments of inertia of the moon about its centre, we have

$$\frac{C - A}{C} = 5\frac{M}{m}\frac{a^3}{c^3} = 0{\cdot}0000375 \qquad \text{.........(13),}$$

$$\frac{C - B}{C} = \frac{5}{4}\frac{M}{m}\frac{a^3}{c^3} = 0{\cdot}0000094 \qquad \text{.........(14),}$$

where the numerical evaluations are based on current knowledge of the moon's mass and size. In addition we have identically $\dfrac{B - A}{C - A} = \dfrac{3}{4}$.

12·71. Now the two ratios $(C - A)/C$ and $(B - A)/C$ for the moon are capable of being found from observation. The former is proportional to the mean inclination of the moon's equator to the ecliptic, and $(B - A)/C$ to the amplitude of the moon's true libration in longitude. Both these propositions rest on purely dynamical considerations*. The inclination can be determined with much certainty; there seems little room for doubt that the corresponding value of $(C - A)/C$ is close to 0·0006289. The amplitude of the libration in longitude, however, has been the subject of much discussion. The balance of opinion at present seems to follow Hayn† in making the ratio $(B - A)/(C - A)$ nearly equal to its theoretical value 0·75; but several investigators have obtained values in the neighbourhood of 0·5, and the difference appears to be attributable only to observational errors‡. In either case both $(C - A)/C$ and $(C - B)/C$ very much exceed their theoretical values.

The discrepancy was first noticed by Laplace, who was content to attribute the high values of these quantities to distortions developed during solidification. There is no reason on this theory why they should have attained any particular magnitudes; the actual values must be regarded as accidental. Let us see whether it is possible to reconcile the data on the basis of known laws.

In the first place, the numerical coefficients depend on the hypothesis that the moon is homogeneous. If we suppose the moon to be composed

* Routh, *Advanced Rigid Dynamics*, 1905, Chap. 12; Tisserand, *Mécanique Céleste*, 2, Chap. 28.

† *Abhand. d. k. Sächs. Gesell. d. Wiss. zu Leipzig*, 30, 1907, 1–103.

‡ Stratton, *Memoirs of R.A.S.* 59, 1909, 257–290.

of two constituents of the same densities as Wiechert's two layers in the earth, the value of the radius of the inner sphere being determined to fit the actual mean density of the moon, it is found that little change is made. The coefficients are multiplied by 0·9; there is no other change, and the discrepancy is slightly increased.

The present determinations of the masses of the earth and moon, and of the mean radius of the moon, are incapable of serious error, and their values in the past cannot have varied much, at least since the moon solidified.

There remains the mean distance of the moon from the earth. Let us suppose, then, that the moon last adjusted itself to the hydrostatic form when considerably nearer the earth than it is now, at a distance c_1, say, and consider the moon's subsequent development as it receded from the earth and the disturbing influences correspondingly diminished. Since, by hypothesis, no further adjustment by plasticity is taking place, the only change of form is that due to pure elastic deformation, and is much less than the change that would take place in a fluid body. Let β_0 denote the value that $(C - A)/C$ would sink to if the moon receded to an infinite distance from the earth without any hydrostatic adjustment taking place. At any time later than the last adjustment, $(C - A)/C$ is composed of two parts, β_0 and the part due to elastic strain. Suppose the latter to be κ times the value of $(C - A)/C$ for a fluid moon at the same distance. Then the conditions at the last hydrostatic adjustment were such that β_0, and the elastic value of $(C - A)/C$ appropriate to distance c_1, together made up the hydrostatic value for the same distance. Now, allowing for the influence of heterogeneity, we have for the hydrostatic value,

$$\frac{C - A}{C} = 4{\cdot}5 \frac{M}{m} \frac{a^3}{c^3} \qquad \text{.........(1)},$$

and our condition is
$$\beta_0 + 4{\cdot}5\kappa \frac{M}{m} \frac{a^3}{c_1{}^3} = 4{\cdot}5 \frac{M}{m} \frac{a^3}{c_1{}^3} \qquad \text{.........(2)}.$$

We see next that the elastic deformation is produced by the contemporary influence of the earth's gravity and the moon's rotation together, and therefore the protuberance always points *exactly* to the earth. Thus the attraction of the earth on this protuberance passes exactly through the centre of the moon, cannot affect the moon's rotation, and thus does not take part in determining the inclination of the moon's axis of rotation. The latter is determined by β_0 alone, and therefore the value of $(C - A)/C$ determined from observations is simply β_0. Thus (2) can be written

$$4{\cdot}5\,(1 - \kappa) \frac{M}{m} \frac{a^3}{c_1{}^3} = 0{\cdot}0006289 \qquad \text{.........(3)}.$$

Now κ is small, and can be shown theoretically to be about 0·013. Also

$$\frac{M}{m} = 81{\cdot}2,$$

$$\frac{a}{c_0} = \frac{1}{220},$$

where c_0 is the moon's present distance. Hence

$$\frac{c_1}{c_0} = 0{\cdot}376.$$

The data can therefore be reconciled by supposing that the moon last adjusted itself to the hydrostatic state when its distance from the earth was about 140,000 km., and its period of revolution $27{\cdot}3\,(0{\cdot}376)^{\frac{3}{2}} = 6{\cdot}3$ of our present days.

It will be seen that the arguments about the variation in $(C - A)/C$ since the last hydrostatic adjustment apply equally well to $(B - A)/C$; therefore the inference that the ratio of these quantities should be 0·75 remains unaltered. If, however, the moon during solidification was executing a libration in longitude of amplitude 40°, this would no longer hold, and the ratio would be reduced to 0·50*.

12·72. The theory just developed requires that the matter within the moon should now depart appreciably from the hydrostatic state, and should have maintained this departure for a long time geologically. It must therefore have a finite strength. The amount of this strength can be evaluated roughly, by using a result of Sir G. H. Darwin†. The greatest stress-difference in a homogeneous spheroid due to a second order zonal inequality over the surface is at the centre, and its magnitude is $\frac{32}{95} g\rho a e$, where g is surface gravity, ρ the density, a the radius of the spheroid, and e its ellipticity. For the moon, with our adopted value of β_0, which would be equal to e if the moon were symmetrical, this would make the stress-difference equal to 2×10^7 dynes/cm.² Since the difference between the moments B and C of the moon is only a quarter of $C - A$, a theory regarding the moon as affected only by a zonal harmonic about its x axis will not be far wrong.

This result is of interest in relation to the imperfections of isostasy on the earth. The quantities $C - A$ and $B - A$ for the moon arise from the moments of inertia, and therefore certainly from uncompensated inequalities of figure. If we adopt the theory given here, they represent a fossil tide that has persisted through geological time. Even if we do not, they still imply a stress-difference of the order of 2×10^7 dynes/cm.² in the moon's interior. We can evade the conclusion that this has existed for a large fraction of the moon's age only by assuming the ellipticities to have been produced recently, and we should still have to explain how or why a solid sphere could spontaneously deform itself throughout its mass so as to produce a general ellipticity of figure of just such a character and in just such a position as to deceive us and leave no observable free oscillations. It hardly seems worth while to attempt this at present; a direct presumption is established that the moon's material can support per-

* Jeffreys, *Mem. R.A.S.* 60, 1915, 187–217.
† *Scientific Papers*, 2, 474–481.

manently a stress-difference of at least 2×10^7 dynes/cm.^{2} Now the anomalies of gravity left unexplained in discussions of isostasy have been seen to point to a strength within the asthenosphere of 4×10^7 to 10^8 dynes/cm.^{2} The widespread gravity anomalies within the continents used in Heiskanen's estimate of the ellipticity of the equator imply a strength of order 6×10^7 dynes/cm.^{2} The same applies to the widespread anomalies of gravity at sea. All these estimates are of the same order of magnitude.

If we had only the terrestrial evidence it might be possible to maintain that the strengths indicated were correct only for stresses lasting for intervals comparable with the time elapsed since the last epoch of mountain formation, say 30 million years, and that much smaller stresses would produce permanent deformation if maintained for a longer time. Such a view has been held by various writers for vitreous solids and even for crystals, though it has neither theoretical nor experimental support. But since on any view seriously held the moon's composition and original thermal state must have closely resembled that of the earth's rocky shell, and the same processes of cooling from the surface and upward concentration of radioactive matter must have determined its thermal history, we cannot assume a fundamental difference in properties between the earth's shell and the moon. The maintenance of the moon's ellipticity of figure therefore provides strong reason to believe that the strength of the asthenosphere inferred from the three types of discussions of gravity is not only genuine but permanent under stresses lasting for times comparable with the age of the earth.

A consequence of the acceptance of the finite strength of the rocky shell is that the assumption of a hydrostatic state within the asthenosphere is only an approximation, and it is of no use to push it beyond the limits of accuracy imposed by the outstanding residuals. Second degree terms in the ellipticity of figure are of little interest. The widespread anomalies over continents and oceans imply that the ellipticity varies from one meridian to another as 1/294 to 1/299, but until they are known in sufficient number to permit interpolation and spherical harmonic analysis not much can be gained by attempting to improve present estimates of the variation of gravity with latitude and of the mean ellipticity. Other harmonics in the form of the geoid, including that expressed by the ellipticity of the equator, cannot be satisfactorily distinguished by present data.

12·8. *Summary*. A theory of the figure of the ocean surface, without reference to the state of the solid interior, has been developed. The approximation based on the assumption of a hydrostatic state for the interior is developed, and it is shown that the adoption of a correct value of the moment of inertia will imply correct values of the precessional constant, the ellipticity, and the variation of gravity over the ocean. This value would be consistent with a modification of Wiechert's hypothesis, so that

the central iron core has a mean density of about 12, and the rocky shell one of 4·3. The differences between these and the densities of iron and dunite at the surface can be legitimately attributed to compression. Gravity surveys have also disclosed small, but definite departures from isostasy over wide areas, which are too large to be attributed to any compensated disturbance, and imply stresses within the asthenosphere comparable with those implied by the unexplained residuals of the theory of isostasy. The observed large differences between the moments of inertia of the moon are explicable on the hypothesis that it solidified when much nearer the earth than it is now, and has maintained the figure it had then; but on this or any other reasonable hypothesis of the maintenance of these differences the moon's interior must have a strength comparable with that inferred for the earth's asthenosphere, and its permanence in the moon gives strong reason to believe that it is also permanent within the earth.

CHAPTER XIII

The Variation of Latitude and the Bodily Tide

"O, that this too too solid flesh would melt,
Thaw and resolve itself into a dew!"
SHAKESPEARE, *Hamlet*.

13·1. There are two phenomena that give information about the elastic yielding of the earth as a whole to disturbing forces. Although seismology and the figure of the earth together give very detailed information about the elasticity within the earth, they do not pronounce definitely as to whether the central core is a true liquid or merely an imperfectly elastic solid that admits distortional waves but absorbs them before they have penetrated through it. These two other phenomena, however, provide the required additional datum and enable us to say that the core is liquid.

Consider a sphere acted on by a bodily force derivable from a gravitational potential $K_n r^n S_n$, where r is the distance from the centre, n a positive integer, S_n a surface harmonic of degree n, and K_n a function of r. The equations of motion are those of 9 (20). If the period of the disturbance is long compared with that of any of the earth's elastic vibrations we can omit d^2u/dt^2; and if the disturbance is isothermal we can put $\gamma = 0$. The earth being heterogeneous, λ, μ, and ρ_0 are functions of r. Putting

$$r^n S_n = W_n \qquad \text{.........(1)}$$

we can assume*

$$u = F\frac{\partial W_n}{\partial x} + GxW_n;\quad \frac{1}{\rho_0}\left\{\frac{\partial}{\partial x}(\rho_0 u) + \frac{\partial}{\partial y}(\rho_0 v) + \frac{\partial}{\partial z}(\rho_0 w)\right\} = -\frac{\rho_1}{\rho_0} = -\phi W_n \qquad \text{.........(2)},$$

where F, G, and ϕ are functions of r. The radial displacement

$$= \frac{1}{r}(xu + yv + zw) = \frac{1}{r}(nFW_n + Gr^2 W_n)$$

$$= \frac{q}{r} W_n \qquad \text{.........(3)},$$

where q is another function of r. Also

$$\phi W_n = -\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z}\right) - \frac{1}{\rho_0}\left(u\frac{\partial \rho_0}{\partial x} + v\frac{\partial \rho_0}{\partial y} + w\frac{\partial \rho_0}{\partial z}\right)$$

$$= -\left(\frac{n}{r}\frac{dF}{dr} + (n+3)\,G + r\frac{dG}{dr}\right) W_n - \frac{1}{r\rho_0}\frac{d\rho_0}{dr} qW_n \qquad \text{.........(4)}.$$

* Love, *Geodynamics*, 1911, 127.

The gravitation potential is of two parts, one arising from an external disturbance, which is supposed known, and the other from the changes of density within the body itself. These give

$$\nabla^2 (K_n W_n) = -4\pi f \rho_1 = -4\pi f \phi \rho_0 W_n$$
$$= \left(\frac{2n}{r}\frac{dK_n}{dr} + \nabla^2 K_n\right) W_n \qquad \text{.........(5).}$$

On substituting in the three equations of motion we find that all break up into two parts, the first for instance involving the sum of a function of r multiplied by $\partial W_n/\partial x$, the other a function of r multiplied by xW_n. The three equations are then satisfied if these two functions of r vanish, and thus we have two differential equations for F and G. The solution for a heterogeneous sphere is very difficult, and has hitherto been achieved only for simplified cases. But we can see at once that the equations involve W_n only through n; if a set of values of F and G has been obtained for one spherical harmonic disturbance, it is applicable to every other disturbing potential of the same degree.

13·11. *Love's numbers.* If the original disturbing potential is $k_2 r^2 S_2$, a solid spherical harmonic of degree 2, we define an 'equilibrium tide' as equal to $k_2 a^2 S_2/g$, where g is undisturbed gravity at the surface. The radial displacement is $aq(a)\, k_2 S_2$. The ratio

$$gq(a)/a = h \qquad \text{.........(1)}$$

specifies the ratio of the actual elevation of the surface to the conventional equilibrium elevation. Also the total gravitational potential amounts at $r = a$ to $K_2(a)\, a^2 S_n$, and we put

$$K_2(a) = (1 + k)\, k_2 \qquad \text{.........(2).}$$

The potential at external points due to the deformation is $kk_2 a^5 S_2/r^3$, and amounts at $r = a$ to k times the disturbing potential. The numbers h and k specify the surface elevation and disturbance of potential due to any tidal potential of the second degree.

13·12. Direct measurement of the displacement of the surface is difficult, but the elevation of the ocean tide on an equilibrium theory is the total gravitation potential divided by g. The height is therefore $(1 + k)\, k_2 a^2 S_2/g$. But the ground is lifted a distance $hk_2 a^2 S_2/g$, and all we can observe is the displacement of sea level relative to the ground, namely $(1 + k - h)\, k_2 a^2 S_2/g$. The ratio of this observed tide to the equilibrium tide is $1 + k - h$. If we know the tidal potential due to the moon, we can calculate the equilibrium tide, and comparison with observation determines $1 + k - h$. It is necessary for accurate results that the observed tides should satisfy an 'equilibrium theory,' that is, that inertia should be negligible, or that the free period should be short compared to the actual period of the tide. The chief results obtained have been those of

Michelson and Gale*, from the disturbance of the levels of water in two vertical tubes with a long connexion underground; and of Schweydar† and Hecker, using horizontal pendulums. Proudman has called attention to the possibility of using the tides in the Red Sea‡. The determination is appreciably affected by the attraction of the tides in the ocean as a whole, which do not satisfy the conditions for the validity of an equilibrium theory. For this reason the determination of Michelson and Gale, being made at a great distance from the nearest sea, is probably the most accurate. It gives $1 - h + k = 0{\cdot}69$.

13·13. The number k is intimately connected with the motion of the earth known as the 'polhode motion' or the 'Eulerian nutation.' Let us consider how a deformable rotating body will move if its rotation is slightly disturbed. The centre of mass is taken as origin, and axes of (x, y, z) are taken through it, turning with an angular velocity whose components are $(\theta_1, \theta_2, \theta_3)$. If the earth was rigid and these axes fixed in it, the velocity in space of any point of it would be

$$(-y\theta_3 + z\theta_2, \ -z\theta_1 + x\theta_3, \ -x\theta_2 + y\theta_1).$$

If now m be the mass of a typical particle of the body and (u, v, w) its velocity components in space, we choose $(\theta_1, \theta_2, \theta_3)$ so as to make the velocities of points rigidly attached to the axes as good an approximation as possible to the actual velocities; that is, we make

$$\Sigma m \{(u - z\theta_2 + y\theta_3)^2 + (v - x\theta_3 + z\theta_1)^2 + (w - y\theta_1 + x\theta_2)^2\}$$

a minimum for variations in $\theta_1, \theta_2, \theta_3$. Hence we have

$$\Sigma m \{z (v - x\theta_3 + z\theta_1) - y (w - y\theta_1 + x\theta_2)\} = 0 \quad \text{.........(1)},$$

with two symmetrical relations. But

$$\Sigma m (yw - zv) = h_1 \quad \text{.........(2)},$$

the angular momentum of the body about the x axis. Also if A, B, C, F, G, H are the moments and products of inertia we have

$$\Sigma m (y^2 + z^2) = A; \ \Sigma mzx = G; \ \Sigma mxy = H \quad \text{.........(3)},$$

and (1) is equivalent to

$$h_1 = A\theta_1 - H\theta_2 - G\theta_3 \quad \text{.........(4)},$$

with two symmetrical relations. These have the same form as for a rigid body§. The equations of motion have the usual form

$$\frac{dh_1}{dt} - h_2\theta_3 + h_3\theta_2 = L \quad \text{.........(5)},$$

where L, M, N are the applied couples, in this case zero. In our problem rotation is almost wholly about the axis of z. $F, G, H, \theta_1, \theta_2$, and the variable

* *Journ. Geol.* **27**, 1919, 585–601.

† *Veröff. d. Zentralbureau d. Int. Erdmessung*, 1921.

‡ *Observatory*, 1925, 386–388.

§ Cf. Lamb, *Higher Mechanics*, 1920, 171–173.

parts of A, B, C are supposed small quantities of the first order. Neglecting their products we can write

$$\left.\begin{aligned} A\frac{d\theta_1}{dt} + (C-B)\,\theta_2\theta_3 - \theta_3\frac{dG}{dt} + F\theta_3{}^2 &= 0 \\ B\frac{d\theta_2}{dt} - (C-A)\,\theta_1\theta_3 - \theta_3\frac{dF}{dt} - G\theta_3{}^2 &= 0 \\ C\frac{d\theta_3}{dt} \qquad\qquad\qquad\qquad &= 0 \end{aligned}\right\} \quad \text{.........(6).}$$

From the last we can put $\theta_3 = n$, a constant, save for small perturbations. Also the undisturbed values of A and B are equal. The direction cosines of the instantaneous axis are, omitting second order terms,

$$(l, m, 1) = (\theta_1/n, \theta_2/n, 1) \quad \text{.........(7),}$$

whence

$$\left.\begin{aligned} A\dot{l} + (C-A)\,nm - \dot{G} + Fn &= 0 \\ B\dot{m} - (C-A)\,nl - \dot{F} - Gn &= 0 \end{aligned}\right\} \quad \text{.........(8).}$$

Now the earth is rotating about its instantaneous axis, and the displacements produced by rotation are equivalent to those given by a disturbing potential per unit mass

$$\tfrac{1}{2}n^2\{(x^2+y^2+z^2) - (lx+my+z)^2\} = \tfrac{1}{2}n^2(x^2+y^2) - n^2z\,(lx+my) \text{...(9)}$$

approximately. The first term is independent of the time; it has been already treated in discussing the figure of the earth. The second is a harmonic of degree 2. It produces an elastic deformation of the earth such as to give an extra external gravitation potential

$$-\,kn^2z\,(lx+my)\,a^5/r^5 \quad \text{.........(10).}$$

But the gravitation potential due to the deformed earth is

$$f\left[\frac{M}{r} + \frac{(A+B+C)\,r^2 - 3\,(Ax^2+By^2+Cz^2-2Fyz-2Gzx-2Hxy)}{2r^5}\right] \quad \text{.........(11)}$$

by MacCullagh's formula; whence

$$3fF = -\,kn^2ma^5;\quad 3fG = -\,kn^2la^5 \quad \text{.........(12),}$$

and

$$\left.\begin{aligned} \left(A + \frac{kn^2a^5}{3f}\right)\dot{l} + \left(C - A - \frac{kn^2a^5}{3f}\right)nm &= 0 \\ \left(A + \frac{kn^2a^5}{3f}\right)\dot{m} - \left(C - A - \frac{kn^2a^5}{3f}\right)nl &= 0 \end{aligned}\right\} \quad \text{.........(13).}$$

If now we introduce a complex variable

$$w = l + \iota m \quad \text{.........(14),}$$

and put

$$\frac{A + kn^2a^5/3f}{C - A - kn^2a^5/3f} = \tau \quad \text{.........(15),}$$

the equations combine into

$$\frac{dw}{dt} - \frac{\iota nw}{\tau} = 0 \quad \text{.........(16),}$$

whence

$$l = \alpha\cos n\,(t-t_0)/\tau;\quad m = \alpha\sin\,(t-t_0)/\tau \quad \text{.........(17),}$$

where a and t_0 are real arbitrary constants, showing that the axis revolves in a circular cone relative to the earth's axis of figure in period τ sidereal days. We have also

$$g = \frac{fM}{a^2} \qquad \text{.........(18)}$$

and can put
$$\mu = n^2 a/g \qquad \text{.........(19)}$$

(the m of Chapter XII). Then

$$\frac{kn^2a^5}{3f} = \frac{k\mu g a^4}{3f} = \frac{k}{3}\mu M a^2 \qquad \text{.........(20).}$$

This is small in comparison with A, and we have nearly

$$\frac{C - A - \frac{1}{3}k\mu Ma^2}{A} = \frac{1}{\tau} \qquad \text{.........(21).}$$

If we put $A/(C - A) = \tau_0$ and use 12·22 (9), we find

$$k = \left(\frac{2\epsilon}{\mu} - 1\right)\left(1 - \frac{\tau_0}{\tau}\right) \qquad \text{.........(22),}$$

a result given by Love and Larmor*.

13·14. For a perfectly rigid earth h and k would be zero, and by 13·13 (22) the period of the motion is τ_0 times the period of rotation. The motion is in fact the polhode motion discovered theoretically by Euler and observable as a wobble whenever a disc or ring is thrown with a spin not quite about its axis of symmetry. The ratio $A/(C - A)$ is known from the rate of the precession of the equinoxes to be 305. Consequently astronomers were led to look for an oscillation of the pole with a period of ten months, and it was only after repeated failures to find a motion with such a period that S. C. Chandler in 1891 found a small variation, apparently consisting of two parts, with periods of a year and 14 months respectively. Both had amplitudes of the order of a tenth of a second of arc. It was then pointed out by Newcomb that elasticity would lengthen the period, and that the 14-monthly motion was really the Eulerian nutation with its period so modified†.

The question may be looked at in another way. In an elastic earth the ellipticity may be considered as composed of two parts. One is due to strain and is always symmetrical about the axis of rotation, and can therefore do nothing to shift this axis within the body. The other is permanent, and affects the rotation just as in a rigid body. It is the latter part, the part of $(C - A)/A$ not due to elastic strain, that determines the period of the motion; and conversely the observed lengthening from 10 to 14 months gives a measure of the elastic strain.

The displacements of the pole are detected through their effects on the observed latitudes of observatories. The latitude of a station is by definition the mean of the altitudes of a circumpolar star when it crosses

* *Proc. Roy. Soc.* A. **82**, 1909, 80, 94.

† *M.N.R.A.S.* **52**, 1892, 336–341.

the meridian above and below the pole*. The altitude is the complement of the zenith distance, and the zenith is in the direction opposite to local gravity. The mean of the altitudes is the altitude of the celestial pole. The colatitude of the station is therefore the angle between local gravity and the earth's instantaneous axis of rotation. A motion of the latter within the earth therefore affects the observed latitudes of all stations. Systematic observations of the variation at six stations in as nearly as possible the same latitude gave the displacements of the pole in two perpendicular directions at intervals of 0·1 year for many years and made its detailed study possible. In addition to this international programme, work of great accuracy has been done at Greenwich.

A full harmonic analysis of the observations has recently been carried out by L. W. Pollak†. From 1890 to 1923 it seemed impossible to improve on a period of 1·20 year for the free vibration. This gives

$$k = 0{\cdot}287 \qquad \text{.........(1).}$$

Combining this with

$$1 - h + k = 0{\cdot}69 \qquad \text{.........(2),}$$

we find

$$h = 0{\cdot}60 \qquad \text{.........(3).}$$

But these values of h and k include effects due to the ocean, which yields completely to disturbing forces. If h and k are to be used to infer properties of the earth's interior it is desirable that the effects of the known fluid layer should as far as possible be removed first. Unfortunately tidal theory is not yet sufficiently developed to permit an accurate estimate of the corrections. According to Larmor‡ the correction to k is about $-0{\cdot}02$, giving

$$k = 0{\cdot}27 \qquad \text{.........(4).}$$

The effect on h is more doubtful, but Street§ has shown that a uniform ocean would increase $1 - h + k$ by 0·04 of itself. This must be reduced to allow for the fact that the ocean covers only part of the earth, which reduces its effect directly and also prevents the ellipsoidal deformations from reaching their full development. We may take the corrected value of $1 - h + k$ to be 0·67, giving

$$h = 0{\cdot}60 \qquad \text{.........(5),}$$

as before.

13·2. The problem of the deformation of a homogeneous incom-

* Refraction is great at the lower passage, and makes this method of finding the latitude inaccurate in practice; the practical method is that of Talcott, depending on observations of different stars near the zenith, and is discussed in *Publ. of R. Observatory, Greenwich*, 'Observations with the Cookson floating telescope, 1911–18.'

† *Gerlands Beiträge*, **16**, 1927, 108–194.

‡ *Proc. Lond. Math. Soc.* **14**, 1915, 440–449.

§ *M.N.R.A.S. Geoph. Suppl.* **1**, 1925, 292–306.

pressible elastic sphere by body forces was solved by Kelvin*. The result is equivalent to

$$h = \frac{5}{2}\frac{2g\rho a}{19\mu + 2g\rho a} \qquad \text{.........(1)},$$

where μ is the rigidity, a the radius, ρ the density, and g surface gravity. With $h = 0{\cdot}6$, $g = 981$ cm./sec.^{2}, $\rho = 5{\cdot}5$ gm./cm.^{2}, and $a = 6{\cdot}4 \times 10^8$ cm., this gives

$$\mu = 1{\cdot}15 \times 10^{12} \text{ dynes/cm.}^2 \qquad \text{.........(2)},$$

rather more than the rigidity of steel.

At the time when Kelvin wrote, this rigidity seemed remarkably large. But it can now be checked by the seismological evidence. The density just below the intermediate layer and the velocity of distortional waves being taken as 3·3 and $4{\cdot}35 \times 10^5$ in c.g.s. units, the rigidity is $6{\cdot}2 \times 10^{11}$ dynes/cm.^{2} But within the rocky shell these values rise to about 5 and $7{\cdot}5 \times 10^5$, giving a rigidity near the base of this shell of $2{\cdot}8 \times 10^{12}$ dynes/cm.^{2} Further, a rigidity of the latter order is correct for over 1000 km. of the rocky shell. The average rigidity within the shell is therefore distinctly larger than that of a homogeneous earth that would fit the tidal data. Further, the assumption of homogeneity puts the mass too far from the centre, where it would be more acted on by tidal forces, so that effectively Kelvin's solution gives the deforming forces a better chance than they have in the actual earth. A smaller rigidity than his would be needed to fit the tidal data in the actual earth; a larger one is needed to fit the seismic evidence.

Formal solutions of problems allowing for departures of the actual earth from Kelvin's conditions have been given by Herglotz, Schweydar, and Love. Herglotz found the deformation of an incompressible sphere of Wiechert's type, with a homogeneous shell over a homogeneous core. Schweydar used parabolic density and rigidity distributions of Roche's type. Love allowed for compressibility in a homogeneous sphere. Herglotz's solution has been applied by Schweydar, Love, and myself to various sets of numerical data. As the effect of compressibility appears to be small, Herglotz's allowance† for heterogeneity probably makes his solution closely applicable to the facts. As a specimen of the results obtained, I may quote my own solution‡ based on the original Wiechert hypothesis with densities 3·2 and 8·2; the rigidities of the shell and core needed to fit the variation of latitude would be $3{\cdot}5 \times 10^{11}$ and $16{\cdot}5 \times 10^{11}$ dynes/cm.^{2} Stoneley§, combining Wiechert's densities with Knott's velocities, allowing for compressibility, and using a numerical method of solution, got a calculated yielding about two-thirds of that observed. All the investigations of tidal phenomena agree in implying much lower rigidities than seismology

* *Phil. Trans.* A, **153**, 1863.

† *Zs. f. Math. u. Phys.* **52**, 1905, 275–299.

‡ *Memoirs R.A.S.* **60**, 1915, 198.

§ *M.N.R.A.S. Geoph. Suppl.* **1**, 1926, 356–359.

does. A way out of the difficulty, however, is opened by the suggestion that the central core, whose radius is about half that of the earth as a whole, is liquid. The densities adopted for the shell and core being those of the modified Wiechert distribution considered in 12·45, with a rigidity of the shell of $1{\cdot}695 \times 10^{12}$ dynes/cm.^{2}, it is found* that zero rigidity of the core would make $h = 0{\cdot}667$, $k = 0{\cdot}372$; a rigidity of the core equal to that of the shell would make $h = 0{\cdot}364$, $k = 0{\cdot}191$. The latter results are much too low to fit the tidal data, and definitely imply that the core is less rigid than the shell absolutely, and not merely in comparison with its density. The results for a liquid core are a little too high, but it appears probable that they would be appreciably reduced if we allowed for variations of density and rigidity within the two layers. A liquid core would then make it possible to reconcile the tidal and seismological data; a core with a rigidity related to its bulk-modulus (about 10^{13} dynes/cm.^{2}) in any ratio admissible for a solid would be entirely impossible.

13·3. We can now proceed to consider the forced variation of latitude due to periodic variations of the products of inertia, that is, to parts of F and G not determined by 13·13 (12). It is convenient to introduce an 'axis of inertia,' whose extremity is the 'pole of inertia,' with direction cosines $(\lambda, \mu, 1)$ defined by

$$\lambda = -\frac{G_1\tau}{A}; \quad \mu = -\frac{F_1\tau}{A} \qquad \text{.........(1)},$$

where F_1 and G_1 are the extra parts of F and G.

This axis is the axis of maximum moment of inertia of the body whose moments and products of inertia are

$$A, A, A\left(1+\frac{1}{\tau}\right), F_1, G_1, H.$$

Using this axis makes it possible to reduce 13·13 (8) to the simple form

$$\left.\begin{aligned} \frac{\tau}{n}\dot{l} + m &= \mu \\ \frac{\tau}{n}\dot{m} - l &= -\lambda \end{aligned}\right\} \qquad \text{.........(2)}.$$

Now l and m may be found from the polar motion, being indeed the component angular displacements of the pole. If then some explanation of the annual variation of latitude is suggested, and we can find the corresponding annual variation of F_1 and G_1, we can compare the two sides of equations (2) and obtain a quantitative test of the hypothesis.

13·31. We have now to express F_1 and G_1, and hence λ and μ, in terms of the mass-transference that occurs during the year. Suppose the annual

* Jeffreys, *M.N.R.A.S. Geoph. Suppl.* **1**, 1926, 371–383.

component of the mass per unit area on an element of the earth's surface to be $\frac{1}{2}\sigma \cos \odot + \frac{1}{2}\sigma' \sin \odot$, where $\odot$ is the sun's longitude. Then

$$F_1 = \iiint \rho yz\,dx\,dy\,dz \text{ through the volume}$$

$$= \iint (\tfrac{1}{2}\sigma \cos \odot + \tfrac{1}{2}\sigma' \sin \odot)\, yz\,dS \text{ over the surface}$$

$$= a^4 \int_0^{\pi}\int_0^{2\pi} (\tfrac{1}{2}\sigma \cos \odot + \tfrac{1}{2}\sigma' \sin \odot) \sin^2\theta \cos\theta \sin\phi\, d\theta\, d\phi \quad \text{...(1)},$$

$$G_1 = a^4 \int_0^{\pi}\int_0^{2\pi} (\tfrac{1}{2}\sigma \cos \odot + \tfrac{1}{2}\sigma' \sin \odot) \sin^2\theta \cos\theta \cos\phi\, d\theta\, d\phi \quad \text{...(2)},$$

where a is the radius of the earth, θ the colatitude, and ϕ the longitude east of Greenwich. The axis of x is thus taken in the plane of the equator and intersecting the meridian of Greenwich, and that of y 90° east of it. σ and σ' are functions of θ and ϕ.

These values of F_1 and G_1, however, require to be corrected for the effect of elastic yielding in the earth under the pressure of the superficial matter. If the variable part of the mass per unit area is m, let m be supposed expanded in spherical harmonics. Let one term in this expansion be $\alpha \sin\theta \cos\theta \sin\phi$, $= \alpha S_2$, say. Then F_1 has exactly the same value as if m were actually equal to αS_2. Now the strain due to any harmonic in m can be treated separately, and the radial displacement will be proportional to the same harmonic as the surface density. Hence no harmonic other than S_2 can contribute anything to F_1. All we need to find, then, is the strain in the earth due to a surface layer of density αS_2. The theory of the elastic deformation of the earth has not yet been carried far enough to find this; all we can say is that the strain will reduce F_1 and G_1 in a constant ratio, which we shall call κ.

Using $A = \frac{1}{3}Ma^2$, where M is the mass of the earth, we have

$$\lambda = \kappa\lambda_0; \quad \mu = \kappa\mu_0; \qquad \text{.........(3)},$$

where

$$\left.\begin{aligned} \lambda_0 &= -\,9''\!\cdot\!1 \times 10^{-3} \int_0^{\pi}\int_0^{2\pi} (\sigma \cos \odot + \sigma' \sin \odot) \sin^2\theta \cos\theta \cos\phi\, d\theta\, d\phi \\ \mu_0 &= -\,9''\!\cdot\!1 \times 10^{-3} \int_0^{\pi}\int_0^{2\pi} (\sigma \cos \odot + \sigma' \sin \odot) \sin^2\theta \cos\theta \sin\phi\, d\theta\, d\phi \end{aligned}\right\} \text{...(4)},$$

where σ and σ' must now be expressed in grams per square centimetre.

We are now in a position to treat various putative causes separately. There are several obvious annual changes in the distribution of mass over the surface that could affect the products of inertia. Evidently, if the whole of the surface of the earth were of similar character, either land or sea, or even if the distribution of land and sea were symmetrical about the polar axis, F_1 and G_1 would necessarily be zero, and there could be no displacement of the axis of rotation. The chief methods of redistribution of matter over the surface in the course of a year are apparently

(1) the seasonal variation in the distribution of air over the surface, shown by the variation of atmospheric pressure observed at the surface;

(2) precipitation of snow, which accumulates in places throughout the winter;

(3) periodical changes in vegetation, such as the formation of deciduous parts of trees, the rise of sap in trees, and the formation of annual parts of herbs.

13·32. The numerical data for the estimation of the effects of the annual redistribution of air have been found from the annual variation of pressure over the earth, given in Bartholomew's *Meteorological Atlas*; though a recalculation with more modern data may be desirable. Allowance must be made for the effect of the pressure on the sea surface in redistributing the water of the ocean. If m be taken to refer to the mass of air and water together per unit surface, we have from the law of indestructibility of matter

$$\iint m ds = 0 \qquad \text{.........(1)}$$

taken over the whole surface of the earth, both land and water.

But for an annual motion the elevation of the ocean surface must have practically its equilibrium value. In other words, the water will adjust itself so that the pressure is uniform over any level surface and therefore has no tendency to produce horizontal movement. In these circumstances the mass per unit area of the surface above a given level must be uniform over the ocean; and the periodical variation of the mass per unit area is therefore the same all over the ocean. Let its value be m'. Then

$$m' \iint dS + \iint m dS = 0 \qquad \text{.........(2)}$$

by (1); where the first integral extends over the sea and the second over the land. If, then, m_1 be the mean value of m over the land,

$$\begin{aligned} m' &= - m_1 \text{ (area of land)/(area of sea)} \\ &= - \cdot 40 m_1 \qquad \text{.........(3).} \end{aligned}$$

Next,

$$\begin{aligned} F_1 &= \iint myz dS \text{ over the whole surface} \\ &= \iint m'yz dS \text{ over the ocean} \\ &+ \iint myz dS \text{ over the land} \\ &= \iint (m + 0{\cdot}40 m_1)\, yz dS \text{ over the land} \qquad \text{.........(4).} \end{aligned}$$

Use has here been made of the fact that $\iint yz dS$, taken over the whole surface, is zero. Similarly

$$G_1 = \iint (m + 0{\cdot}40 m_1)\, zx dS \text{ over the land} \qquad \text{.........(5).}$$

On carrying out the numerical integrations, we find

$$\left.\begin{aligned} \lambda_0 &= 0''{\cdot}0040 \cos \odot + 0''{\cdot}0051 \sin \odot \\ \mu_0 &= - 0''{\cdot}0134 \cos \odot + 0''{\cdot}0659 \sin \odot \end{aligned}\right\} \qquad \text{.........(6).}$$

A correction is, however, required; or rather, a correction already applied needs to be removed. The pressures given in the meteorological charts are not the true atmospheric pressures at the places concerned, but these

pressures reduced to sea-level according to a formula given by Laplace*, which depends in part upon the temperature. To find the mass of air over a given area, however, we need, not these modified values, but the true local pressures; the correction included in the charts therefore needs to be taken out again. This requires that to the values of λ_0 and μ_0 just found we must add†

$$\left.\begin{aligned} \lambda_0 &= 0''{\cdot}0000 \cos \odot - 0''{\cdot}0014 \sin \odot \\ \mu_0 &= 0''{\cdot}0128 \cos \odot - 0''{\cdot}0315 \sin \odot \end{aligned}\right\} \quad \ldots\ldots\ldots(7).$$

The atmosphere and ocean together therefore contribute an amount

$$\left.\begin{aligned} \lambda_0 &= 0''{\cdot}0040 \cos \odot + 0''{\cdot}0037 \sin \odot \\ \mu_0 &= -0''{\cdot}0006 \cos \odot + 0''{\cdot}0344 \sin \odot \end{aligned}\right\} \quad \ldots\ldots\ldots(8).$$

The greater part of the contribution comes from the seasonal change in Central Asia, whose chief meteorological correlate is the monsoons. For this reason the greatest part of the displacement of the pole of inertia is towards Central Asia in summer, when the amount of air there is a minimum, and away from it in winter.

It may be remarked that it would be incorrect to say that the part of the annual variation of latitude derived from (8) is due to the annual variation of pressure. It has been seen that the annual variation of latitude depends simply on the products of inertia of the earth as a whole, the atmosphere being included, and the atmospheric pressure is used only as an observable quantity that enables us to infer the mass per unit surface; that it does not affect the motion directly is seen from the facts that it is equal to the product of gravity into the mass per unit area, and that gravity nowhere occurs in the results.

13·33. *The Effect of Precipitation.* In this investigation we are concerned only with those modes of transport of matter that cause a variation in the mass per unit area over the surface. For instance, if rain falls upon a particular spot, and at once runs away to the sea, it contributes nothing to the elements of the products of inertia corresponding to that point. Similarly, if it evaporates at once, it contributes nothing to the elements derived from the solid and liquid parts of the earth; the presence of the water vapour in the atmosphere, however, will affect the mass of the atmosphere locally, and consequently its mass will be included in the variations already inferred from the distribution of atmospheric pressure. Thus the effect of water vapour in the atmosphere has already been considered. This section is concerned only with those parts of the precipitation that remain on the ground for a considerable time. These are, first, the water that is absorbed by the soil; second, the snow that lies on the

* *Mécanique Céleste*, edition of 1880, 4, 294.

† The numerical data for this computation and that of (6) are in *M.N.R.A.S.* **76**, 1916, 499–525.

ground in winter. Data for the first part are almost entirely lacking. In this country the soil is wetter in winter than in summer, principally owing to the reduced evaporation; the annual variation of rainfall is not nearly so great as that of soil-moisture. In countries, however, where most of the rain falls in summer, this state of affairs may be reversed. On the whole it seems probable that the effect of soil-moisture is small in comparison with that of snowfall.

A rough estimate of the effect of snowfall can be made easily. Snow accumulates in winter to a considerable depth over a great part of Canada and the United States, and over the greater part of Asia down to about the latitude of the Tian Shan. In South America the amount of snowfall is small, and in Australia it is negligible. On the assumption that in the areas of Asia and North America where snow accumulates, it does so at a uniform rate until the spring thaw, I estimate* that the contributions to λ_0 and μ_0 are

$$\lambda_0 = -0''\cdot0238 \cos \odot + 0''\cdot0175 \sin \odot,$$

$$\mu_0 = -0''\cdot0126 \cos \odot + 0''\cdot0093 \sin \odot.$$

13·34. The effect of vegetation is small compared with that of the annual displacements of air. In woodland and natural grassland the mass of vegetation per unit area in summer exceeds that in winter by one or two grams per square centimetre, whereas the corresponding difference for atmospheric motions in Central Asia is about 10 grams per sq. cm. Thus the effect of vegetation is much less than that of atmospheric displacements; but it is sufficiently great to be observable, were it not associated with other larger displacements of the same period. The contributions from this source are roughly

$$\lambda_0 = 0''\cdot0017 \cos \odot - 0''\cdot0038 \sin \odot,$$

$$\mu_0 = 0''\cdot0024 \cos \odot - 0''\cdot0055 \sin \odot.$$

13·35. Accumulation of ice and snow round the poles may be summed up under two types.

The first includes the deposition of snow during the winter, with its removal partly by evaporation and partly by glacier flow during the summer. This part would correctly be included under snowfall. On account, however, of the symmetry of the permanently frozen areas about the poles, it seems very unlikely that this portion can contribute anything appreciable to the movement of the instantaneous axis.

The second type is due to the freezing of the oceans. Freezing in mid-ocean would necessarily lead to no change in the mass per unit area; for the ice must float in such a way that its mass is equal to that of the water displaced, and it is also necessarily equal to that of the water it was formed from. Thus the water displaced has the same volume as the

* *Loc. cit.* p. 520.

frozen water, and the freezing produces no vertical movements of water such as are necessary to start horizontal motion. Ice attached to the margin of a continent may behave differently, but even in this case, on account of symmetry about the polar axis, the contribution to the motion of the axis must be small.

13·36. Summing up our results, we have the following:

	λ_0	μ_0
Atmosphere and ocean	$0''{\cdot}0037 \sin \odot + 0''{\cdot}0040 \cos \odot$	$0''{\cdot}0344 \sin \odot - 0''{\cdot}0006 \cos \odot$
Snowfall	$0''{\cdot}0175 \sin \odot - 0''{\cdot}0238 \cos \odot$	$0''{\cdot}0093 \sin \odot - 0''{\cdot}0126 \cos \odot$
Vegetation	$-0''{\cdot}0038 \sin \odot + 0''{\cdot}0017 \cos \odot$	$-0''{\cdot}0055 \sin \odot + 0''{\cdot}0024 \cos \odot$
Total	$0''{\cdot}0174 \sin \odot - 0''{\cdot}0181 \cos \odot$	$0''{\cdot}0382 \sin \odot - 0''{\cdot}0108 \cos \odot$

The actual motion of the pole of rotation was found by Kimura* to be given by

$$\left.\begin{aligned} l &= 0''{\cdot}011 \sin \odot - 0''{\cdot}096 \cos \odot \\ m &= -0''{\cdot}051 \sin \odot - 0''{\cdot}004 \cos \odot \end{aligned}\right\} \text{from } 1893{\cdot}8 \text{ to } 1899{\cdot}8,$$

and
$$\left.\begin{aligned} l &= 0''{\cdot}008 \sin \odot - 0''{\cdot}064 \cos \odot \\ m &= -0''{\cdot}056 \sin \odot + 0''{\cdot}006 \cos \odot \end{aligned}\right\} \text{from } 1900{\cdot}0 \text{ to } 1907{\cdot}0,$$

and by the present writer† to be

$$\left.\begin{aligned} l &= 0''{\cdot}038 \sin \odot - 0''{\cdot}055 \cos \odot \\ m &= -0''{\cdot}061 \sin \odot - 0''{\cdot}017 \cos \odot \end{aligned}\right\} \text{from } 1907{\cdot}1 \text{ to } 1914{\cdot}0.$$

Taking $\tau = 430$, and using 13·3 (2) with $\kappa = \frac{3}{4}$, we derive from these

$$\left.\begin{aligned} \lambda &= 0''{\cdot}006 \sin \odot - 0''{\cdot}036 \cos \odot \\ \mu &= 0''{\cdot}062 \sin \odot + 0''{\cdot}009 \cos \odot \end{aligned}\right\} \text{from } 1893{\cdot}8 \text{ to } 1899{\cdot}8,$$

$$\left.\begin{aligned} \lambda &= 0''{\cdot}015 \sin \odot + 0''{\cdot}002 \cos \odot \\ \mu &= 0''{\cdot}019 \sin \odot + 0''{\cdot}015 \cos \odot \end{aligned}\right\} \text{from } 1900{\cdot}0 \text{ to } 1907{\cdot}0,$$

$$\left.\begin{aligned} \lambda &= 0''{\cdot}018 \sin \odot - 0''{\cdot}017 \cos \odot \\ \mu &= 0''{\cdot}004 \sin \odot + 0''{\cdot}028 \cos \odot \end{aligned}\right\} \text{from } 1907{\cdot}1 \text{ to } 1914{\cdot}0.$$

The most striking feature of these motions of the pole of inertia inferred from the motion of the pole of rotation is their variability among themselves. They resemble one another and the calculated motion in that all are of the same order of magnitude and retrograde (i.e. from east to west), and the largest term in the motion inferred from meteorological data, that in μ_0 involving $\sin \odot$, is twice represented by the largest term in the values inferred from astronomy; but that is all that can be said. The variation is probably too great to be attributed to observational error, and certainly too great to be attributed to real variations in meteorological conditions from year to year.

* *Ast. Nach.* **181**, 4344. His x is l, and his y is $-m$.

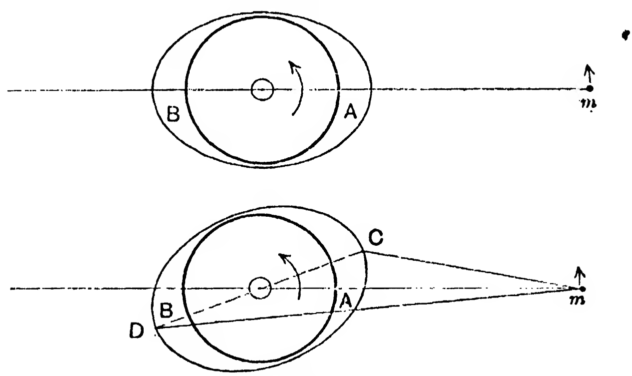
† *Loc. cit.* pp. 523–525.

CHAPTER XIV

Tidal Friction

"Ye fyul, scz a chep, it's a bonny myun,
They've ketched and myed it a clock fyece."
Tyneside song, 'The Fiery Clock Fyece.'

14·1. *Nature of Tidal Friction.* In our discussion of the resonance theory of the origin of the moon, it was seen that this theory implies that the moon was initially within a few thousand kilometres of the earth's surface, and that the earth at that time rotated in a few hours. In these circumstances the angular momentum of the earth's rotation much exceeded that attributable to the moon's revolution. On the other hand, the angular momentum of the moon's revolution is now about five times that of the earth's rotation, and it was assumed as essential to the resonance theory that some physical process exists that has been competent to produce the requisite reduction in the speed of rotation of the earth and the associated increase in the mean distance of the moon.



Figs. 15 and 16.

Tidal friction is qualitatively, and probably quantitatively, capable of producing such an effect. The nature of the action of the tides in slowing up the earth's rotation and driving the moon further away may be illustrated as follows. Let m in Figs. 15 and 16 denote the moon, while the circle enclosed by the thick line represents the equatorial section of the solid earth. The arrows indicate the directions of rotation and revolution. If the solid earth were a perfect sphere, devoid of friction, the tides raised by the moon in a deep ocean would be at A and B, vertically below the moon and straight opposite to it, as shown by the continuous ellipse in Fig. 15. Moderate rotation would affect the height of the tides, but not their position with reference to the moon. During the revolution of a particular point P

on the earth's surface, the water level above it will still rise and fall when rotation is taken into account, the maximum tide height occurring when the point is just below or just opposite to the moon, and the minima at the two positions of quadrature. If, however, friction is present, it will delay the times of maximum and minimum elevation at a particular station, in accordance with the well-known effect of friction in making small oscillations lag in phase. Thus the highest tide at P will not occur till some time after it has passed A or B, and the form of the section of the water level by the equator will resemble the ellipse in Fig. 16, the high tides being at C and D.

Now let us consider the attraction of the tides on the moon. For simplicity the two high tides may be replaced by two heavy particles at the opposite points C and D. The attraction of C on the moon is along mC, and that of D is along mD. That of C is the greater, for it is nearer to m. Neither force acts accurately along the radius joining the centre of the earth to the moon, and therefore both have components at right angles to it. That arising from C is evidently the greater, both because the resultant force due to C is the greater, and because the angle CmO is greater than DmO. Thus there will be on the whole a force on the moon with a component in the direction of its revolution.

Similarly we may consider the attraction of the moon on the two tidal protuberances. We see that the force on C is greater than that on D, and further, on account of the fact that the angle CmO is greater than DmO, its line of action passes further from the centre of the earth. On both grounds it has a greater moment about the centre of the earth. Therefore an effect of the moon's attraction on the tides is to produce a couple tending to turn the earth in the direction opposite to its rotation, and thus to slow up its rotation.

Similar results will evidently hold even if there is no ocean, but the interior of the earth is imperfectly elastic. On account of its elasticity, the form of the solid earth will be somewhat distorted by the moon's tidal action; and imperfect elasticity will cause the greatest elevations to be, not at A and B, but at places like C and D. Thus bodily tides, like oceanic tides, will tend to accelerate the moon and retard the earth.

14·2. *Effects of Tidal Friction.* The above simple discussion does not consider the complications introduced by the irregular form of the oceans. It is not to be expected, however, that such a complication will prevent the existence of the couples. The solar tides, further, behave similarly to those raised by the moon, and therefore have to be taken into account in a quantitative discussion of the effects of tidal friction.

Let the masses of the earth, moon and sun be respectively M, m, and m_1. Let the mean angular velocities of the moon and sun about the earth be n and n_1, and their distances from the earth c and c_1. Let the retarding

couples acting on the earth due to the lunar and solar tides be respectively $-N$ and $-N_1$. Let the earth's angular velocity of rotation be Ω, and its principal moment of inertia about its polar axis C. Then

$$n^2c^3 = f(M+m);\ n_1^2c_1^3 = f(m_1+M) \qquad \text{.........(1)},$$

where f is the constant of gravitation. Put

$$c = c_0\xi^2,\ n = n_0\xi^{-3};\ c_1 = c_{10}\xi_1^2,\ n_1 = n_{10}\xi_1^{-3} \qquad \text{.........(2)},$$

where the zero suffixes indicate the present values of the quantities they are attached to.

The moon revolves about the centre of gravity of the earth and moon together, the distance of which from the moon is $Mc/(M+m)$. The angular velocity of the moon being n, the linear velocity is $Mcn/(M+m)$, and therefore the angular momentum of the moon about the centre of the earth is $Mc^2n/(M+m)$. That of the earth's rotation is $C\Omega$. By Newton's Third Law, to the couple $-N$ acting on the earth must correspond a force acting on the moon, whose moment about the centre of the earth is $+N$. The solar tides will have no secular effect on the moon, for their period is different, and therefore in the course of a lunar synodic month they are presented to the moon in all aspects. Thus they will accelerate the moon in one half of the month as much as they retard it in the other half. Thus the solar tides will not affect the moon; similarly the lunar tides, in the long run, will not affect the sun. Now noticing that $c^2n = c_0^2n_0\xi$, we see by taking moments about the centre of the earth for the moon, sun, and earth respectively that

$$\frac{Mm}{M+m}c_0^2n_0\frac{d\xi}{dt} = N \qquad \text{.........(3)},$$

$$\frac{m_1M}{m_1+M}c_{10}^2n_{10}\frac{d\xi_1}{dt} = N_1 \qquad \text{.........(4)},$$

$$C\frac{d\Omega}{dt} = -N - N_1 \qquad \text{.........(5)}.$$

If E be the total mechanical energy in the system, it decreases at a rate equal to the sum of the rates of performance of work by the angular motions in overcoming the operating couples. This gives

$$-\frac{dE}{dt} = (N+N_1)\Omega - Nn - N_1n_1 \qquad \text{.........(6)}.$$

Consider the effects of the lunar tides separately. These give

$$-\frac{dE}{dt} = N(\Omega - n) \qquad \text{.........(7)}.$$

Since the couples arise from dissipation of energy, the left side is essentially positive. Thus N has the same sign as $\Omega - n$. In the earth-moon system this is positive, and therefore N is positive. This argument is independent of any assumption about the form of the ocean or the nature of the elasticity of the earth. Similarly we can see that N_1 must be positive. It follows from (3) and (4) that the mean distances of the moon and sun must

be increasing, and accordingly their mean motions, in comparison with an inertial system, must be decreasing. It follows again from (5) that the rate of rotation of the earth must be decreasing.

14·21. Consider now what the ratio of N to N_1 is likely to be. The potential at the surface of a body due to a mass m at distance c can be put in the form

$$\frac{fm}{R} = \frac{fm}{c}\left(1 - \frac{x}{c} + \frac{2x^2 - y^2 - z^2}{2c^2}\right) \qquad \text{.........(1)},$$

where the axis of x joins the centres of the two bodies, that of z is the polar axis, and that of y is perpendicular to both. The acceleration of the body as a whole has components

$$\left(-\frac{fm}{c^2}, 0, 0\right) = -\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right)\frac{fmx}{c^2} \qquad \text{.........(2)}.$$

The distortion of the body is due to the difference between the actual potential and the potential that would displace it without changing its form. Apart from a constant term, this difference is equal to

$$\frac{fm}{2c^3}(2x^2 - y^2 - z^2) \qquad \text{.........(3)}.$$

We may now introduce spherical polar coordinates at the centre of the earth, so that

$$\left.\begin{aligned} x &= r\sin\theta\cos\phi \\ y &= r\sin\theta\sin\phi \\ z &= r\cos\theta \end{aligned}\right\} \qquad \text{.........(4)}.$$

Then θ is the colatitude of the point considered, and ϕ is the difference between its longitude and that of the moon. (3) becomes

$$\frac{fmr^2}{2c^3}\{(\tfrac{1}{2} - \tfrac{3}{2}\cos^2\theta) + \tfrac{3}{2}\sin^2\theta\cos 2\phi\} \qquad \text{.........(5)}.$$

The two terms of this are spherical harmonics. The first is independent of the longitude of the point considered, and therefore can produce only a permanent tide not moving with the moon. The second term generates the semidiurnal tide, for we see that it has maxima where $\phi = 0$ and $\phi = \pi$; that is, just below and just opposite to the moon. Thus the tide-generating potential is

$$U = \frac{3}{4}\frac{fmr^2}{c^3}\sin^2\theta\cos 2\phi \qquad \text{.........(6)}.$$

The tide raised by the moon will be expressible as the resultant of a number of surface harmonic displacements. The couple due to the attraction of the moon on an element of the earth's mass is $\rho\dfrac{\partial U}{\partial\phi}d\tau$, where $d\tau$ is an element of volume and ρ the local density. The total couple is therefore

$$-N = -\iiint \frac{3}{2}\frac{fmr^2\rho}{c^3}\sin^2\theta\sin 2\phi\, r^2\,dr\,d\omega \qquad \text{.........(7)},$$

where $d\omega$ is an elementary solid angle, so that

$$d\tau = r^2 dr\, d\omega \qquad \text{.........(8).}$$

The integral is to be taken throughout the earth. If we consider only the oceanic tide, ρ is a function of r and θ only within the solid earth, and therefore the integral up to the mean position of the ocean surface contributes nothing to the couple. If the height of the oceanic tide is h, and the radius of the earth is A, the integral up to the actual ocean surface becomes

$$-N = -\iint \frac{3fmA^4}{2c^3} \rho h \sin^2\theta \sin 2\phi\, d\omega \qquad \text{.........(9),}$$

where ρ now refers to the ocean alone. We see that none of the harmonics involved in h, save that in $\sin^2\theta \sin 2\phi$, can contribute anything to the couple. Let this one be expressible in the form $H \sin^2\theta \cos 2(\phi - \epsilon)$. Evidently 2ϵ is the phase lag introduced by tidal friction, corresponding to double the angle COA in Fig. 16. The portion of this harmonic involving $\cos 2\phi \cos 2\epsilon$ vanishes on integration with regard to ϕ, and we are left with

$$\begin{aligned} -N &= -\frac{3fmA^4}{2c^3} \rho H \int_0^\pi \int_0^{2\pi} \sin^5\theta \sin^2 2\phi \sin 2\epsilon\, d\theta\, d\phi \\ &= -\frac{8}{5}\pi \frac{fmA^4}{c^3} \rho H \sin 2\epsilon \qquad \text{.........(10).} \end{aligned}$$

A similar discussion is applicable to the solar tides.

If suffix 1 indicates that the quantity it is attached to refers to the sun or to the solar tide, we evidently have from (10)

$$N/N_1 = \frac{mH}{c^3} \sin 2\epsilon \Big/ \frac{m_1 H_1}{c_1^3} \sin 2\epsilon_1 \qquad \text{.........(11).}$$

We notice that the ratio $\frac{m}{c^3}\Big/\frac{m_1}{c_1^3}$ is the ratio of the amplitudes of the two tide-generating potentials due to the moon and sun. If then the periods of the two tides were equal, and if the laws determining the tides were linear in the displacements, we could calculate the two tides separately; their phase lags would then be equal, and their amplitudes would be in the ratio of those of the disturbing potentials. Thus we should have

$$\frac{N}{N_1} = \left(\frac{m/c^3}{m_1/c_1^3}\right)^2 \qquad \text{.........(12).}$$

This ratio is about 5·1.

14·22. We readily see that the above argument can be generalized so as to be applicable to the bodily tide. The displacements at all points of the interior will be nearly in proportion in the lunar and solar tides, provided only that the periods are far from resonance. This hypothesis is justified by the fact that the period of free vibration of a fluid earth

would be less than two hours, and this would be shortened by elasticity; while the periods of both tides are about 12 hours. Thus internal motion will introduce into (9) and (10) only such factors as will be the same for both moon and sun, and therefore (11) will hold. But the conditions for inferring (12) from (11) are also satisfied, for we cannot suppose that the small strains involved in bodily tides produce friction proportional to any power but the first. Thus in this case

$$N/N_1 = 5{\cdot}1.$$

14·23. Lags in the ocean tides produced by viscosity give the same ratio. Much tidal friction, however, takes the form of skin friction in regions of strong tidal currents; it will indeed be seen later that this is probably the most important type operative at the present time. Skin friction is proportional to the square of the velocity of the tidal current, and therefore it is no longer justifiable to treat the lunar and solar tides separately and add the results. If, however, friction is too small to alter the character of the tides seriously, we may estimate the ratio in another way. The effect of friction is to introduce a reaction from the sea where it occurs upon the water of the ocean as a whole. Friction changes sign with the tidal current, and the reaction will therefore contain a portion with a period equal to that of the tide; this portion will produce a small tide in the ocean, out of phase with the tide unaffected by friction. The tide unaffected by friction gives by hypothesis no dissipation of energy and therefore no tidal couples; these arise from the attraction of the moon and sun on the small tides produced by the reaction caused by friction.

In a region where friction is considered, let u be the velocity of the lunar current, and v that of the solar current. Suppose the origin of time so chosen that we can write

$$u = a \sin pt; \quad v = b \sin qt \qquad \text{.........(1).}$$

Evidently b is considerably less than a, and their ratio is approximately the same all over the earth. Further,

$$p = 2(\Omega - n); \quad q = 2(\Omega - n_1) \qquad \text{.........(2).}$$

To a first approximation only we may suppose the whole velocity to be $a \sin pt + b \sin qt$. We require to use this so as to proceed to an approximation to the form of the frictional forces. The frictional force is $F = k(u + v)^2$, directed against the resultant velocity; or, in a form that puts in evidence the fact that it changes sign with the velocity,

$$F = k\,|\,u + v\,|\,(u + v) \qquad \text{.........(3),}$$

where $|\,u + v\,|$ denotes the absolute value of $u + v$.

Let the two parts of F whose periods are equal to those of the semidiurnal tides be $P \sin pt$ and $Q \sin qt$. Other parts can give only tides whose periods are different from those of the disturbing potentials, and

will therefore give only periodic effects on the moon and sun. We need therefore consider only these two terms. Now

$$\pi P = \operatorname*{Lim}_{n\to\infty} \frac{1}{n}\int_0^{2n\pi} F \sin pt\, dpt \qquad \text{.........(4)},$$

$$\pi Q = \operatorname*{Lim}_{n\to\infty} \frac{1}{n}\int_0^{2n\pi} F \sin qt\, dqt$$

$$= \frac{q}{p}\operatorname*{Lim}_{n\to\infty} \frac{1}{n}\int_0^{2n\pi} F \sin qt\, dpt \qquad \text{.........(5)}.$$

Put

$$pt = \theta,\ q/p = r \qquad \text{.........(6)}.$$

Then

$$\pi P = \operatorname{Lim} \frac{1}{n}\int_0^{2n\pi} k \,|\, a\sin\theta + b\sin r\theta \,|\, (a\sin\theta + b\sin r\theta)\sin\theta\, d\theta \quad \text{...(7)},$$

$$\pi Q = \operatorname{Lim} \frac{q}{pn}\int_0^{2n\pi} k \,|\, a\sin\theta + b\sin r\theta \,|\, (a\sin\theta + b\sin r\theta)\sin r\theta\, d\theta \quad \text{...(8)}.$$

Neglecting b in (7) we find

$$\pi P = 2ka^2\int_0^{\pi} \sin^3\theta\, d\theta = \tfrac{8}{3}ka^2 \qquad \text{.........(9)}.$$

Let θ_m be the mth zero of $(a\sin\theta + b\sin r\theta)$, not counting the origin. If we neglect b^2, we see easily that

$$\theta_m = m\pi - (-1)^m \frac{b}{a}\sin rm\pi \qquad \text{.........(10)},$$

$$\pi Q = \operatorname{Lim} \frac{kq}{pn}\sum_{m=0}^{2n-1}(-1)^m\int_{\theta_m}^{\theta_{m+1}} (a\sin\theta + b\sin r\theta)^2 \sin r\theta\, d\theta \quad \text{......(11)}.$$

The general term in the sum in (11) is

$$\tfrac{1}{2}(-1)^{m-1}\left[\begin{array}{l} a^2\left\{\frac{1}{r}\cos r\theta - \frac{1}{2(2+r)}\cos(2+r)\theta + \frac{1}{2(2-r)}\cos(2-r)\theta\right\} \\ + 2ab\left\{\cos\theta - \frac{1}{2(2r+1)}\cos(2r+1)\theta + \frac{1}{2(2r-1)}\cos(2r-1)\theta\right\} \\ + \text{terms in } b^2 \end{array}\right]_{\theta_m}^{\theta_{m+1}}$$

$$= \tfrac{1}{2}(-1)^{m-1}\left[a^2\frac{1}{r}\{\cos r(m+1)\pi - \cos rm\pi\}\right.$$

$$- \frac{a^2}{2(2+r)}\{\cos(2+r)(m+1)\pi - \cos(2+r)m\pi\}$$

$$\left. + \frac{a^2}{2(2-r)}\{\cos(2-r)(m+1)\pi - \cos(2-r)m\pi\}\right]$$

$$+ \tfrac{1}{4}ab\{\cos 2r(m+1)\pi - \cos 2rm\pi\}$$

$$- \tfrac{1}{8}ab\{\cos 2(1+r)(m+1)\pi - \cos 2(1+r)m\pi\}$$

$$- \tfrac{1}{8}ab\{\cos 2(1-r)(m+1)\pi - \cos 2(1-r)m\pi\}$$

$$+ 2ab\left[1 - \frac{(-1)^{m-1}}{4(2r+1)}\{\cos(2r+1)(m+1)\pi - \cos(2r+1)m\pi\}\right.$$

$$\left. + \frac{(-1)^{m-1}}{4(2r-1)}\{\cos(2r-1)(m+1)\pi - \cos(2r-1)m\pi\}\right] \qquad \text{.........(12)}.$$

These terms are to be added for all values of m from zero to $2n-1$. Evidently the terms with m in the argument form several series of cosines of angles in arithmetical progression, and therefore their sum never exceeds a definite finite amount. Thus the integral gives

$$\pi Q = \lim_{n\to\infty} \frac{kq}{pn} 2ab\,(2n+\lambda),$$

where λ does not tend to infinity with n

$$= 4kqab/p \qquad \text{.........(13).}$$

Thus
$$\frac{Q}{P} = \frac{3}{2}\frac{q}{p}\frac{b}{a} \qquad \text{.........(14).}$$

This gives us the required approximation to the ratio of the reactions. Now since by hypothesis the ocean is far from resonance, and r is not far from unity, the amplitudes of the tides introduced by friction will be in proportion to the corresponding terms in the force that produces them, and the phase lags will be equal. Thus in 14·21 (11) we have

$$\frac{H}{H_1} = \frac{P}{Q} = \frac{2}{3}\frac{pa}{qb} \qquad \text{.........(15).}$$

The factor q/p is nearly unity, and may be omitted. The ratio b/a arises from the theory before friction was taken into account; hence in calculating it the assumption that the laws involved are linear in the displacements is correct, and therefore

$$\frac{a}{b} = \frac{m/c^3}{m_1/c_1^3} \qquad \text{.........(16).}$$

Thus (15) gives
$$\frac{H}{H_1} = \frac{2}{3}\frac{m/c^3}{m_1/c_1^3} \qquad \text{.........(17),}$$

and finally substituting in 14·21 (11) we have

$$\frac{N}{N_1} = \frac{2}{3}\left(\frac{m/c^3}{m_1/c_1^3}\right)^2 \qquad \text{.........(18)}$$
$$= 3{\cdot}4.$$

Thus the solar tidal friction is a larger fraction of the lunar in this case than when the friction follows a linear law.

14·3. *The Secular Accelerations of the Sun and Moon.* Let us now consider the effect of the variations in the angular motions on observations of astronomical phenomena. Suppose that n, n_1 and Ω are known from observations over a short interval near time zero, and that from the values found we infer that a known fixed star would cross the meridian of Greenwich at time T if all these quantities were constant. Then the effect of the variation in the speed of the rotation of the earth is to put the earth ahead in time T by an angle $\frac{1}{2}T^2\frac{d\Omega}{dt}$, if we neglect variations in $\frac{d\Omega}{dt}$. The earth's angular velocity being Ω, the meridian of Greenwich will therefore reach a given star in time $T - \frac{1}{2}\frac{T^2}{\Omega}\frac{d\Omega}{dt}$ instead of in time T;

in other words, the time of transit of a given star across the meridian is hastened by $\frac{T^2}{2\Omega}\frac{d\Omega}{dt}$. The moon moves among the stars with angular velocity n, and therefore the alteration of the time of the observation makes the moon an angle $\frac{nT^2}{2\Omega}\frac{d\Omega}{dt}$ behind its calculated position when the star transits. On the other hand, the change in its own angular velocity puts it ahead by an angle $\frac{1}{2}T^2\frac{dn}{dt}$ in time T. Altogether, then, it is $\frac{1}{2}T^2\left(\frac{dn}{dt}-\frac{n}{\Omega}\frac{d\Omega}{dt}\right)$ in front of its calculated position when the star transits. Thus, in comparison with the positions inferred for time T, the moon is subject to an advance through an angular distance $\frac{1}{2}T^2\left(\frac{dn}{dt}-\frac{n}{\Omega}\frac{d\Omega}{dt}\right)$. This is equivalent to saying that the moon appears to have, relative to the stars, a steady secular acceleration $\frac{dn}{dt}-\frac{n}{\Omega}\frac{d\Omega}{dt}$. Let us denote this by ν. The sun similarly will have a secular acceleration, which we shall denote by ν_1. Then

$$\nu=\frac{dn}{dt}-\frac{n}{\Omega}\frac{d\Omega}{dt} \qquad \text{.........(1)},$$

$$\nu_1=\frac{dn_1}{dt}-\frac{n_1}{\Omega}\frac{d\Omega}{dt} \qquad \text{.........(2)}.$$

But from 14·2 (2) $\quad \frac{dn}{dt}=-3n_0\xi^{-4}\frac{d\xi}{dt};\quad \frac{dn_1}{dt}=-3n_{10}\xi_1^{-4}\frac{d\xi_1}{dt} \qquad$ ....(3).

Let us now substitute in (1) and (2) the values of the differential coefficients found in 14·2 (3), (4) and (5). Then

$$\nu=-3\frac{M+m}{Mm}\frac{N\xi^{-4}}{c_0^{\,2}}+\frac{N+N_1}{C\Omega}n_0\xi^{-3} \qquad \text{.........(4)},$$

$$\nu_1=-3\frac{m_1+M}{m_1 M}\frac{N_1\xi_1^{-4}}{c_{10}^{\,2}}+\frac{N+N_1}{C\Omega}n_{10}\xi_1^{-3} \qquad \text{.........(5)}.$$

So long as intervals of only a few thousand years are considered, ξ and ξ_1 may be taken equal to unity without sensible error. Let us put

$$\frac{Mm}{M+m}\frac{c_0^{\,2}n_0}{C\Omega_0}=\kappa \qquad \text{.........(6)},$$

so that κ is the present value of the ratio of the orbital angular momentum to the rotational angular momentum of the earth. Taking

$$C=0{\cdot}334MA^2 \qquad \text{.........(7)},$$

we have
$$\kappa=4{\cdot}82 \qquad \text{.........(8)}.$$

Then (4) reduces to $\quad \nu=\frac{M+m}{Mmc^2}\{(\kappa-3)N+\kappa N_1\} \qquad$ ....(9),

where the zero suffixes have now been dropped, since the quantities are nearly equal to their present values.

The ratio of the first term on the right of (5) to the second is

$$\frac{3N_1}{N+N_1}\frac{c^2 n}{c_1{}^2 n_1}\frac{m}{M+m}\frac{1}{\kappa}.$$

The first factor is of order unity, the second 10^{-4}, the third 10^{-2}, the last $\frac{1}{5}$. This ratio is therefore very small, and the first term on the right of (5) may be neglected. Then (5) becomes

$$\nu_1 = \kappa\frac{M+m}{Mm}\frac{n_1}{n}\frac{N+N_1}{c^2} \qquad \text{.........(10).}$$

14·31. Now ν and ν_1 are capable of being found from a comparison of present-day astronomical observations with ancient ones. In particular, it is evident that a determination of the time of the occultation of a star by the moon, or of the conjunction of the moon with a star, will give ν directly. Again, the moon in time T will gain on the sun in longitude by $\frac{1}{2}(\nu-\nu_1)T^2$ in comparison with the calculated motion, and therefore all lunar eclipses will occur earlier by an interval $\frac{1}{2}\frac{\nu-\nu_1}{n-n_1}T^2$. Thus observations of the times of eclipses make it possible to find $\nu-\nu_1$.

The magnitude of a lunar eclipse is determined by the distance of the moon from the node at the time of conjunction or opposition to the sun. The motion of the node has no part arising from tidal friction, and is therefore known from the purely gravitational theory of the moon's motion. The longitude of the moon at the time of eclipse is increased by an amount $-\frac{1}{2}\frac{n}{n-n_1}(\nu-\nu_1)T^2$ on account of the earlier time of the eclipse, but also by $\frac{1}{2}\nu T^2$ on account of its own secular acceleration. Thus the effect of the two accelerations jointly is to increase the moon's longitude at the time of eclipse by $\frac{1}{2}\frac{n\nu_1-n_1\nu}{n-n_1}T^2$. The magnitudes of the eclipses therefore give $n\nu_1-n_1\nu$. The times and magnitudes together give ν and ν_1 separately.

The same arguments would apply to a solar eclipse as seen from the centre of the earth; but owing to the secular change in the rate of the earth's rotation the spot on the surface where the magnitude of the eclipse is greatest will be altered. Thus the discussion of solar eclipses is more complicated than that of lunar ones.

It may be noticed, incidentally, that

$$n\nu_1 - n_1\nu = n\frac{dn_1}{dt} - n_1\frac{dn}{dt},$$

and therefore the magnitudes of the eclipses do not involve the change in the rotation of the earth. This result is evident *ab initio*, since they involve only the positions of the sun and moon and the centre of the earth. Again, dn_1/dt has been seen to be very small, so that practically the magnitudes of the eclipses give a direct determination of dn/dt.

Observations of the time of passage of the sun across the equator, when

the precession of the equinoxes is known, give the secular acceleration of the sun directly.

Many ancient observations by Greek, Babylonian, Chinese, and Egyptian astronomers have been discussed with a view to the determination of the secular accelerations. The most recent and the most thorough of these discussions is that of Dr J. K. Fotheringham. The following table is taken from his summary* of the information obtainable directly from observations of several different types. His L is equivalent to $\frac{1}{2}\nu$ of this chapter. The unit is a second of arc per century per century.

Lunar eclipses				
Times	Magnitudes	Solar eclipses	Occultations and Conjunctions	Equinoxes
—	—	21·0	$21 \cdot 60 \pm 1 \cdot 40$	—
—	$3 \cdot 56 \pm 0 \cdot 90$	2·0	—	$3 \cdot 66 \pm 0 \cdot 54$
$17 \cdot 8 \pm 2 \cdot 6$	—	19·0	—	—

Fotheringham has discussed the solar eclipses afresh†, and it is shown by the diagram on p. 123 of his paper that a secular acceleration of the sun of $2'' \cdot 1/(\text{century})^2$, and one of the moon of $20'' \cdot 7/(\text{century})^2$ would satisfy them all, but no smaller value of the solar secular acceleration is permissible. It might, however, be as large as $3'' \cdot 3/(\text{century})^2$, combined with one of the moon of $22'' \cdot 0/(\text{century})^2$. He decides that on the whole, judging purely on the observational evidence, the most probable values of the secular accelerations are $21'' \cdot 6/(\text{century})^2$ and $3'' \cdot 0/(\text{century})^2$; but they satisfy the observations more accurately than the probable errors would lead one to expect, and consequently some latitude of interpretation is permissible.

The lunar theory, however, indicates a secular acceleration‡ of the moon of $12'' \cdot 2/(\text{century})^2$. The portion attributable to tidal friction is therefore only the unexplained excess, namely about $9'' \cdot 0/(\text{century})^2$.

14·32. With these values of ν and ν_1 it should be possible to solve 14·3 (9) and (10) for N and N_1 and to find their ratio, and also to substitute in 14·2 (6) and find the rate of dissipation of energy directly, without any hypothesis as to the nature of the tidal friction. It appears, however, that the solar acceleration is uncertain by a larger fraction of its amount than the lunar, and it may be preferable to determine theoretically the ratio of the two accelerations, and thus to infer the solar acceleration from the lunar. Even this cannot, however, be done with much accuracy if the friction varies as the square of the velocity, for squares of the height of the solar tide had to be neglected in 14·23 in inferring the ratio of the frictional couples, and an error of the order of 10 per cent. in the result is therefore to be anticipated.

* *M.N.R.A.S.* **80**, 1920, 578–581.

† *Ibid.* **81**, 1920, 104–126.

‡ There was considerable discussion among dynamical astronomers on this point in the first half of the nineteenth century. It was finally cleared up by J. Couch Adams.

Now 14·3 (9) and (10) give

$$\frac{\nu}{\nu_1} = \frac{\frac{\kappa - 3}{\kappa} N + N_1}{N + N_1} \frac{n}{n_1} \qquad \text{.........(1).}$$

When N/N_1 tends to zero, this tends to $n/n_1 = 13{\cdot}3$. This is obvious without analysis, for it corresponds to the case where all friction is in the solar tides; and then dn/dt is zero, and dn_1/dt insignificant. Thus the whole of the secular accelerations arise from errors in the time introduced by variations in the rate of rotation, and are therefore in the ratio of the mean motions. Then to a secular acceleration of the moon of $9''{\cdot}0/(\text{century})^2$ corresponds one of the sun of $0''{\cdot}7/(\text{century})^2$.

When N/N_1 tends to infinity, so that all the friction is in the lunar tides, ν/ν_1 tends to $\frac{\kappa - 3}{\kappa} \frac{n}{n_1} = 5{\cdot}0$. This gives

$$\nu_1 = 1''{\cdot}8/(\text{century})^2,$$

which must be the maximum value possible of the solar secular acceleration.

With N/N_1 equal to 5·1, we get

$$\nu/\nu_1 = 6{\cdot}3, \quad \nu_1 = 1''{\cdot}44/(\text{century})^2.$$

With N/N_1 equal to 3·4, we have

$$\nu/\nu_1 = 7{\cdot}2, \quad \nu_1 = 1''{\cdot}26/(\text{century})^2.$$

We notice that the largest possible value of the solar secular acceleration is somewhat less than the lowest permitted by Fotheringham's investigations. This suggests that either an unknown cause is producing a secular acceleration of the sun, or that part of the observed value is error*.

We have from 14·3 (9)

$$N \frac{M + m}{Mmc^2} \left\{(\kappa - 3) + \kappa \frac{N_1}{N}\right\} = \nu \qquad \text{.........(2),}$$

and from 14·2 (6) $\quad -\frac{dE}{dt} = N(\Omega - n)\left\{1 + \frac{N_1}{N} \frac{\Omega - n_1}{\Omega - n}\right\} \qquad$ ....(3).

Taking $\quad \nu = 9''{\cdot}0/(\text{century})^2 = 4{\cdot}5 \times 10^{-24}/\text{sec.}^2,$

$$N/N_1 = 5{\cdot}1,$$

we have $\quad \dfrac{N}{C} = \dfrac{73\nu}{1 + \frac{8}{3}\frac{N_1}{N}} = 2{\cdot}2 \times 10^{-22}/\text{sec.}^2 \qquad$ ....(4),

$$-\frac{dE}{dt} = 1{\cdot}5 \times 10^{19} \text{ ergs per second} \qquad \text{.........(5).}$$

* Increase of mass of the sun is not a possible explanation. Cf. *M.N.R.A.S.* **84**, 1924, 533–534.

With the same value of ν, but

$$N/N_1 = 3{\cdot}4,$$

we find
$$\frac{N}{C} = 1{\cdot}9 \times 10^{-22}/\text{sec.}^2 \qquad \text{.........(6)},$$

$$-\frac{dE}{dt} = 1{\cdot}39 \times 10^{19} \text{ ergs per second} \qquad \text{.........(7)}.$$

14·4. *Tidal Friction in the Sea: Quantitative Estimates.* Let us now consider where the dissipation of energy takes place. In the open ocean the two components of horizontal velocity u and v, except near the bottom, satisfy differential equations of the form

$$\frac{du}{dt} - 2\omega v = g\frac{\partial}{\partial x}(\bar{\zeta} - \zeta) \qquad \text{.........(1)},$$

$$\frac{dv}{dt} + 2\omega u = g\frac{\partial}{\partial y}(\bar{\zeta} - \zeta) \qquad \text{.........(2)},$$

where x and y are elements of arc in two perpendicular horizontal directions, ζ is the elevation of the sea-level above its mean position and $\bar{\zeta}$ the height of the equilibrium tide. Also

$$\omega = \Omega \cos\theta \qquad \text{.........(3)}.$$

We shall expect that in the absence of close resonance ζ and $\bar{\zeta}$ will be of the same order. Thus u and v will be of order $g\bar{\zeta}/\omega A$, where A is the radius of the earth. Now $g\bar{\zeta}$ is equal to the tidal disturbing potential, and therefore, by 14·21 (6), its maximum value is $\frac{3}{4}\frac{fmA^2}{c^3}$, making $\bar{\zeta}$ 26 cm. Thus u and v are of order 1 cm./sec. The frictional force is $k\rho\,(u^2 + v^2)$, directed against the resultant velocity, and therefore the rate of dissipation of energy is $k\rho\,(u^2 + v^2)^{\frac{3}{2}}$ per unit area, where $k = 0{\cdot}002$ and ρ is the density of water. Thus the dissipation per square centimetre is of order 0·004 erg/sec. The area of the whole ocean being about $3{\cdot}67 \times 10^{18}$ cm.^{2}, the dissipation as a whole must be of the order of 10^{16} ergs/sec., which is a small fraction of that seen to be required to account for the secular accelerations. Thus the chief part of the tidal dissipation cannot arise from friction in the open ocean.

The validity of the above estimate depends on whether the Osborne Reynolds criterion for turbulent motion is satisfied. If the motion were purely viscous, the kinematic viscosity being ν, and if the influence of the boundary were considerable up to a distance h from it, the criterion for purely viscous motion is that uh/ν shall be less than about 300*. For semidiurnal motions in a viscous medium†

$$h^2 = \nu/\omega,$$

so that
$$uh/\nu = u/(\nu\omega)^{\frac{1}{2}}$$
$$= 800.$$

Thus the criterion for viscous motion is not satisfied, and the motion is turbulent.

* Cf. *Phil. Mag.* **49**, 1925, 793–807.

† Lamb, *Hydrodynamics*, 1924, 586.

14·41. It was, however, assumed explicitly in the last section that the conditions are such that ζ is not great compared with $\bar{\zeta}$. If the form of the bottom is such that the tides are highly magnified by resonance or by the shallow water effect, the argument will not hold. Such circumstances cannot exist in the ocean as a whole, but they may well hold locally. Indeed places where tidal currents have velocities very much greater than the 1 cm./sec. inferred for the open ocean in the last section are known to everybody. In such places the dissipation per unit area must much exceed that in the open ocean, especially since the rate of dissipation per unit area is proportional to the cube of the velocity. The question is, whether the increase in the rate of dissipation per unit area is enough to compensate for the limited area of the regions concerned and make the total dissipation in them exceed that in the open ocean.

The question was answered in the affirmative by G. I. Taylor*, who determined the rate of dissipation of energy in the tides in the Irish Sea in two independent ways. One method is the natural one of estimating the velocities all over the area considered, and hence inferring the dissipation per unit area. On integrating this over the whole region we find the total dissipation.

The alternative method is to find the rate of performance of work on the sea by the ocean and the moon together. This is found to be, on the average, positive. The energy in the sea is, however, not increasing steadily, and therefore the energy supplied from outside must be dissipated as fast as it is received. Thus the rate of dissipation can be found. If we consider a straight line across the sea as the boundary, and take the axis of y along this line, the velocity across it is u. Let D be the depth of the water at any point. The pressure at any depth z below the mean position of the surface is $g\rho\,(z+\zeta)$, since the depth below the actual free surface is $z+\zeta$. Thus the entering water does work at a rate $g\rho\,(z+\zeta)\,u$ per unit area of the section. Integrating this with regard to z from $z=-\zeta$ to $z=D$, supposing u independent of z, we see that the rate of performance of work by the water entering vertically below an element of the boundary of unit length is

$$\int_{-\zeta}^{D} g\rho\,(z+\zeta)\,u\,dz = \tfrac{1}{2}g\rho\,(D+\zeta)^2\,u \qquad \text{.........(1).}$$

In addition the entering water brings in its own energy. Taking the mean sea-level as the zero of potential, we see that the mean potential within a column of unit cross section extending to height ζ and depth D is $-\tfrac{1}{2}g\,(D-\zeta)$. The mass of such a column is $\rho\,(D+\zeta)$. The potential energy is therefore $-\tfrac{1}{2}g\rho\,(D^2-\zeta^2)$. Now the horizontal cross section of the column of water that crosses in unit time unit length of the boundary is u. Thus the potential energy of the entering water enters at a rate

$$-\tfrac{1}{2}g\rho\,(D^2-\zeta^2)\,u \qquad \text{.........(2)}$$

per unit length of the boundary.

* *Phil. Trans.* A 220, 1919, 1–33.

The kinetic energy of the entering water evidently contributes energy to the sea at a rate

$$\tfrac{1}{2}\rho\,(D+\zeta)\,(u^2+v^2)\,u \qquad\qquad \text{.........(3)}$$

per unit length.

Combining these three sources of energy, we find for the whole rate of inflow of energy

$$g\rho u\,(D\zeta+\zeta^2)+\tfrac{1}{2}\rho\,(D+\zeta)\,u\,(u^2+v^2) \qquad\qquad \text{.........(4)}$$

per unit length of the boundary.

Now in general ζ is small compared with D. Also u and v are in general of the order of $c\zeta/D$, where c is the velocity of a tidal wave in water of depth D. But

$$c^2 = gD,$$

so that $D\,(u^2+v^2)$ is of order $g\zeta^2$. Thus much the largest part of the entering energy is contributed by the term $g\rho uD\zeta$. On integration along the boundary the whole rate of transfer of energy across the boundary is seen to be

$$\int g\rho Du\zeta dy \qquad\qquad \text{.........(5)}$$

taken along the boundary from end to end. If this again is integrated with regard to the time through a whole period of the motion, we shall obtain the total inflow of energy across the boundary during a period; and if these results are added for all the boundaries of the sea, the total rate of inflow of energy into the sea may be found.

We require in addition the work done by the moon. The disturbing potential due to the moon being U, the work done by the moon when an element of the ocean surface of area dS is raised through a height $d\zeta$ is $U\rho d\zeta dS$. Putting

$$U = g\bar{\zeta} \qquad\qquad \text{.........(6),}$$

so that $\bar{\zeta}$ is the height of the equilibrium tide, we see that the work done by the moon on the sea in a period is

$$\iint g\rho\bar{\zeta}\,\frac{d\zeta}{dt}\,dtdS \qquad\qquad \text{.........(7),}$$

where the integral with regard to t is to be taken over a period, and that with regard to S over the whole area of the sea. This added to the result of the integration of (5) through a period gives the whole supply of energy to the sea; and this energy must be equal to the dissipation of energy in the sea during a period.

14·411. Taylor found the rate of dissipation of energy in the Irish Sea by these two methods. He found that energy enters through St George's Channel and the North Channel at a mean rate of $6{\cdot}4 \times 10^{17}$ ergs per second. The rate of performance of work on the sea by the moon is, on an average, about $-\,4{\cdot}3 \times 10^{16}$ ergs per second. The total, about 6×10^{17} ergs per second, is absorbed by tidal dissipation. The alternative method, in which the dissipation all over the area is estimated simply

from the velocity distribution, gave $5{\cdot}2 \times 10^{17}$ ergs per second, agreeing with the other within the limits of error of the determinations of velocity used.

The dissipation in the Irish Sea alone is therefore about sixty times what was found for the open ocean as a whole. The conjecture that the greater velocity of the currents may more than make up for the smaller area therefore proves to be correct. The rate of dissipation is indeed 1/20 of that required to account for the whole of the secular acceleration of the moon, and it was therefore suggested by Taylor that the dissipation of energy in the whole of the shallow sèas of the earth may be enough to explain the whole of the secular acceleration.

14·412. It may be noticed, however, that the above estimate for the Irish Sea refers definitely to spring tides, when the tidal currents are at a maximum. To find the mean dissipation throughout the lunar month a correcting factor must therefore be applied. Let us then proceed to estimate this factor. Let θ be the phase of the lunar tide, and $r\theta$ that of the solar tide. Put $r = 1 - s$, so that s is $\frac{1}{25}$. Let A be the amplitude of the lunar current and $A\nu$ that of the solar current. Then the total current is

$$A\{\cos\theta + \nu\cos(1-s)\theta\} = A(1 + 2\nu\cos\theta + \nu^2)^{\frac{1}{2}}\cos\left(\theta - \tan^{-1}\frac{\nu\sin s\theta}{1+\nu\cos s\theta}\right),$$

which is now expressed as a simple harmonic motion with a slowly varying amplitude and period. The amplitude at springs is $A(1+\nu)$. The rate of dissipation is proportional to the cube of the current, and therefore that over a lunar day is proportional to the cube of the amplitude during that day. Thus the ratio of the mean dissipation to the dissipation at springs is the average of

$$(1 + 2\nu\cos s\theta + \nu^2)^{\frac{3}{2}}/(1+\nu)^3.$$

If ν^6 be neglected, the numerator of this is $1 + \frac{9}{4}\nu^2 + \frac{9}{64}\nu^4$. Assuming that the velocities vary in proportion to the vertical ranges, we have

$$\frac{1+\nu}{1-\nu} = 2{\cdot}3,$$

$$\nu = 0{\cdot}39,$$

$$\frac{1 + \frac{9}{4}\nu^2 + \frac{9}{64}\nu^4}{(1+\nu)^3} = 0{\cdot}51.$$

Thus the correcting factor is about $0{\cdot}5$, and the number of Irish Seas required to account for the secular acceleration of the moon is between 40 and 50.

14·413. Taylor's methods have been extended by the present writer*

* *Phil. Trans.* A 221, 1920, 239–264.

to include most of the shallow seas of the globe. The results are as follows. All refer to spring tides. The unit is 10^{18} ergs per second.

European waters:		Asiatic waters:		North American waters:	
Irish Sea	0·6	South China Sea	Small	Hudson Strait	0·2
English Channel	1·1	Yellow Sea	1·1	Hudson Bay	Small
North Sea	0·7	Sea of Okhotsk	0·4	Fox Strait	1·4
		Bering Sea	15·0	Bay of Fundy	0·4
		Malacca Strait	1·1		

The other seas of the globe appear unlikely to contribute much to the dissipation. In Europe, the Mediterranean, Baltic, and White Sea are so narrow at their entrances that little tide can enter, and consequently the tidal currents within them are unable to contribute much to the dissipation. The Bay of Biscay is too deep to magnify the currents much. In Asia, the South China Sea contributes little because the tide in it is almost wholly diurnal; the disturbing potential that produces the diurnal tide arises from the inclination of the moon's orbit to the equator, and dissipation in it contributes nothing to the secular acceleration. Australian waters contribute little, for similar reasons; while the Gulf of Mexico resembles the Mediterranean.

In the case of the Yellow Sea and Malacca Strait it has been possible to apply both of Taylor's methods of calculating the dissipation, and concordant results have been obtained. In the Bay of Fundy the currents vary so much from point to point that only the method based on the inflow of energy is useful. In the others it has been necessary to rely wholly on the estimates based on the velocity. The results, especially those for the Bering Sea, the most important region of all, are therefore subject to revision as more accurate knowledge of the tides and tidal streams becomes available.

The total dissipation is found to be 22×10^{18} ergs per second at spring tides. Applying the correcting factor 0·51 to obtain the average dissipation, we find $1{\cdot}1 \times 10^{19}$ ergs per second, 80 per cent. of what is required to account for the whole of the secular acceleration of the moon. It would by itself give a secular acceleration of $7''$ per century per century.

14·414. The agreement between the dissipation in shallow seas and that necessary to account for the lunar secular acceleration is much closer than the data would entitle us to expect. Two-thirds of that found takes place in the Bering Sea; this determination may be in error by half its amount if the estimates of the currents there are systematically wrong by 25 per cent., a possible contingency. What we are entitled to assert, however, on the evidence before us, is that the dissipation is certainly enough to account for a large fraction of the secular acceleration, and that there is nothing to prove that it is incapable of accounting for the whole of it*.

* A discussion based on substantially the same data, and giving comparable results, is given by W. Heiskanen in *Ann. Acad. Scient. Fennicae*, **18** A, 1921, 1–84. See also W. D. Lambert, *Bull.* **11**, *Section d'Oceanographie du Conseil International de Recherches*, 1928.

In particular, any portion of the secular acceleration can be attributed to bodily tidal friction only as a result of independent proof that bodily friction must produce such an amount; there is no insufficiency in the theory of tidal friction in shallow seas that can justify such a course*.

It is uncertain whether the dissipation in any other coastal regions is notable in comparison with those already considered. The only partially enclosed seas not treated so far are some of those in the North-West Passage. There is an extensive shallow region off the coast of Patagonia, but it is in no way enclosed, being perfectly open to the Atlantic. Thus it is difficult to make any reliable inference about the currents in it. Many records of tidal currents along the coast are given in the *Admiralty Pilots*, but all of them seem to refer to currents up rivers or near their mouths, or to currents over bars and shoals; in either case the general currents must be magnified. There seem to be no data about currents more than a few miles out to sea.

The dissipation over local shallows like shoals and bars, and in narrow bays and straits, has been systematically ignored in this discussion, except where it has been automatically taken into account in the determination of the excess of the entering over the issuing energy. The chief reason is the utter impossibility of finding it from our present data. It can be proved that variations of depth across a channel do not affect the longitudinal currents much, so that shallows long compared with their width will not affect the result; but roughly circular shoals, and shallow regions extending right across a channel, may contribute an appreciable amount. The agreement found in the cases of the Irish Sea, the Yellow Sea, and the Strait of Malacca between the results of the two methods of computing the dissipation shows that at any rate the shoals do not contribute an amount comparable with that in the regions as a whole; for one method systematically includes the effect of shoals, and the other systematically omits it.

The fjords of the west coasts of Norway, Greenland, and North and South America, and the Scottish sea lochs, are innumerable. They are, however, mostly very deep, and strong tidal currents are therefore not to be expected over much of their area. Strong currents do, however, occur locally in them. No attempt has yet been made to compute the dissipation in these regions, but it is probably small.

Along the open shore, again, there must be some dissipation; the strong currents usually do not extend more than a few miles out to sea, but they exist along a very long stretch of coast, and the aggregate dissipation in them may be appreciable.

14·42. *Bodily Tidal Friction.* Let us proceed now to consider the possible effects of bodily imperfection of elasticity. We saw in Chapter XI

* Tidal friction in shallow water due to resistance proportional to the square of the velocity was discussed provisionally by Dr W. D. Macmillan, Carnegie Publication 107, 1909.

that imperfection of elasticity in a solid may be manifested in two ways, either by a slowness of elastic recovery after strain, called elastic afterworking, or by steady deformation under stress, called plasticity. It was seen to be possible for plasticity to appear only when the stress exceeded a certain value, which was called the 'strength'; but it was realized that the strength might be zero. It is possible for elastic afterworking and plasticity to be present in the same substance.

No attempt was made to state quantitative laws of these properties, and it appears probable that in actual substances the laws are very complex. In the case of bodily tides in the earth, however, we can make some progress. The amplitude of the bodily tide is, like that of the ocean tide, of the order of a metre. Thus each part of the earth is stretched during the lunar day by amounts comparable with 10^{-7} of its linear dimensions. It is incredible that if any terms in the friction depend on second and higher powers of the displacements, they can be appreciable when the displacements are as small as this. On the other hand, in solid friction and some forms of plasticity, imperfection of elasticity arises only when the stresses exceed a certain amount, and the frictional stress above this point remains constant. But we can hardly suppose the set-point to correspond to an extension comparable with 10^{-7}, and if it is larger the substance will be perfectly elastic for the small stresses involved in bodily tides.

14·421. If then any imperfection of elasticity is present, we must expect it to depend on the first power of the displacements and their rates of change. It is easy to obtain two plausible forms of imperfection of elasticity that show the characteristic features of plasticity and elastic afterworking respectively. We have seen that if P represents a tangential stress, and E the corresponding distortional strain, the relation between them in a perfectly elastic solid is

$$P = \mu E \qquad \text{.........(1)},$$

where μ is the coefficient of rigidity. Now it may happen that the distortion is changing, and if so, the stress is not only maintaining the distortion at its temporary value, but also is producing new distortion. There may be a further resistance to the latter process; it vanishes with the rate of distortion, by definition, and, since it must be linear in the distortion and its rates of change, must therefore be proportional to the rate of increase of distortion. Thus we must have

$$P = \mu\left(E + \mathfrak{t}_2 \frac{dE}{dt}\right) \qquad \text{.........(2)},$$

where $\mathfrak{t}_2$ is a constant with the dimensions of a time. We see that if the stress is removed the strains diminish exponentially to zero, falling to e^{-1} of their original values in time $\mathfrak{t}_2$. If the system is in a state of steady strain, we shall have

$$P = \mu E$$

as before; thus for motions periodic in times long compared with t_2 the substance will behave as if perfectly elastic. For motions with periods short compared with t_2, on the other hand, the term in $t_2 \dfrac{dE}{dt}$ will much exceed that in E. If then μ is finite, any finite value of E would make the right side of (2) greater than the left. E must therefore be zero. Thus when the motions have short periods, and μ is finite, the substance behaves as if perfectly rigid. Thus this type of imperfection of elasticity implies not weakness, but additional stiffness. For this reason it will be referred to as 'firmoviscosity.' It is evidently a particular case of elastic afterworking.

On the other hand, if μt_2 is finite but μ zero, the relation takes the form

$$P = \nu \frac{dE}{dt},$$

which is the characteristic form of the stress-strain relation in a viscous fluid. If finally μt_2 is also zero, P is necessarily zero, and we have a perfect fluid.

14·422. On the other hand, it is possible that the imperfection of elasticity may be of plastic type; that is, if the stress is kept constant, the strain will increase steadily. Then E will contain two parts. The first, P/μ, arises from simple elastic strain. The second must be supposed to increase at a rate proportional to P. Thus we must have

$$\mu E = P + \frac{1}{t_1}\int P\,dt \qquad \text{.........(1)},$$

where t_1 is a constant with the dimensions of a time. In this case a sudden stress gives an immediate strain P/μ as in a perfectly elastic solid, but the strain then proceeds to increase at a uniform rate $P/\mu t_1$ so long as the stress is kept constant. When the stress is removed, the solid does not return to its original configuration, however long it may be left; it loses the purely elastic strain P/μ and retains a permanent set measured by $PT/\mu t_1$, where T is the time of application of the stress.

This type of imperfection of elasticity is called 'elasticoviscosity.' The behaviour of elasticoviscous solids under periodic stress is fundamentally different from that of firmoviscous ones. If the period is long compared with t_1, the second term in (1) will much exceed the first, and the stress-strain relation may be rewritten

$$\mu t_1 \frac{dE}{dt} = P,$$

which is the form appropriate to a viscous liquid. Thus the elasticoviscous solid approximates to a viscous liquid for any long period forces, whatever its rigidity may be. The firmoviscous solid never approximates to a liquid when the rigidity is finite, and behaves as if perfectly elastic for long-period stresses.

If, on the other hand, the period is short compared with t_1, the second

term in (1) is small, and the elasticoviscous solid behaves as if perfectly elastic. In the corresponding case a firmoviscous solid behaves as if perfectly rigid.

14·423. It is possible to combine both types of imperfection of elasticity in one substance. For, if we have

$$\mu\left(E + \mathfrak{t}_2 \frac{dE}{dt}\right) = P + \frac{1}{\mathfrak{t}_1}\int P dt \qquad \text{.........(1)},$$

the substance will follow the firmoviscous law if $\mathfrak{t}_1$ is infinite, and the elasticoviscous one if $\mathfrak{t}_2$ is zero. Such a substance will flow indefinitely with long-continued stresses, but the partial recovery on release will be gradual, whereas in a purely elasticoviscous substance it is instantaneous. Under quickly changing stresses the material will behave as if perfectly rigid, like a firmoviscous substance.

We see that if any problem of elastic strain has been solved for a perfectly elastic solid, the behaviour of an imperfectly elastic one, so long as squares of the displacements can be neglected, can be inferred by simply writing

$$\mu\left(1 + \mathfrak{t}_2 \frac{d}{dt}\right)\bigg/\left(1 + \frac{1}{\mathfrak{t}_1 (d/dt)}\right) \text{ for } \mu \qquad \text{.........(2)}.$$

In particular, the behaviour of a firmoviscous solid can be found by replacing μ by $\mu\left(1 + \mathfrak{t}_2 \frac{d}{dt}\right)$, and that of an elasticoviscous one by replacing μ by $\mu\bigg/\left(1 + \frac{1}{\mathfrak{t}_1 (d/dt)}\right)$. A body nearly perfectly elastic has $\mathfrak{t}_1$ large and $\mathfrak{t}_2$ small, and (2) is nearly

$$\mu\left(1 + \mathfrak{t}_2 \frac{d}{dt} - \frac{1}{\mathfrak{t}_1 (d/dt)}\right) \qquad \text{.........(3)}.$$

The elasticoviscous law was formulated by J. Clerk Maxwell*, while the firmoviscous one† was suggested to me privately by Sir J. Larmor.

14·424. It may be said at once that there is little positive evidence for any type of viscosity within the earth. The rocky shell behaves as if perfectly elastic, and the central core as a perfect liquid, for most of the known phenomena involving stresses less than 10^7 dynes/cm.^{2} or so.

The damping of the surface waves of earthquakes as they advance suggests some kind of imperfection of elasticity in the upper layers, but part of the damping may be due to scattering. The observed amplitudes of the bodily distortional waves indicate little absorption in transit even when they have almost grazed the central core. If a wave type is given by $\exp \iota\,(pt - \kappa x)$, where p is real, κ is given by

$$(p/\kappa)^2 = \alpha\mu/\rho \qquad \text{.........(1)},$$

where α is 1 for bodily distortional waves and does not differ much from it for surface waves. If then imperfection of elasticity changes μ to

* *Phil. Mag.* 35, 1868, 134. † *M.N.R.A.S.* 77, 1917, 449–456.

$\mu\,(1 + \iota b)$, where b is small, its effect on κ is to multiply it by $1 - \frac{1}{2}\iota b$, and the damping factor for increasing distance is $\exp\,(-\frac{1}{2}b\kappa x)$. But also, by 14·423 (3),

$$\iota b = \mathrm{t}_2 \iota p - \frac{1}{\mathrm{t}_1 \iota p} \qquad \text{.........(2),}$$

that is

$$b = p\mathrm{t}_2 + \frac{1}{p\mathrm{t}_1} \qquad \text{.........(3),}$$

so that the damping factor is

$$\exp\left(-\tfrac{1}{2}\kappa p\mathrm{t}_2 - \tfrac{1}{2}\frac{\kappa}{p\mathrm{t}_1}\right)x \qquad \text{.........(4).}$$

Assuming that the period is 10 s., and the velocity 4 km./sec., and that the amplitude is reduced in the ratio $e : 1$ as the wave advances 6000 km., we find that the data would be satisfied if either

$$\mathrm{t}_1 = 750 \text{ sec.}, \ \mathrm{t}_2 = 0; \text{ or } \mathrm{t}_2 = 0{\cdot}004 \text{ sec.}, \ \mathrm{t}_1 = \infty \qquad \text{.........(5).}$$

14·425. The transmission of compressional waves through the central core without appreciable loss can be shown* to indicate that its kinematic viscosity is under about 2×10^9 cm.²/sec.

14·426. The free variation of latitude has been shown by Pollak to have persisted for about 30 years without any marked systematic change. The period of this motion is given by 13·13 (22), which may be written

$$1 - \frac{\tau_0}{\tau} = k\lambda \qquad \text{.........(1),}$$

where k is Love's number and λ a known constant. With sufficient accuracy we can suppose that k is inversely proportional to the rigidity of the shell; this amounts to supposing the elastic straining of the earth to be determined to a first approximation by rigidity and not by gravity. When viscosity is neglected $1 - \tau_0/\tau$ is real and about $\frac{1}{4}$. Allowing for viscosity and introducing a b as in the last paragraph, we get

$$1 - \frac{\tau_0}{\tau} = \tfrac{1}{4}\,(1 - \iota b) \qquad \text{.........(2),}$$

while $2\pi/\tau = p$, if we temporarily use the day as the unit of time and suppose the displacements proportional to $\exp \iota pt$. Thus

$$\frac{p\tau_0}{2\pi} = \tfrac{3}{4} + \tfrac{1}{4}\iota b \qquad \text{.........(3),}$$

and viscosity multiplies p by $1 + \frac{1}{3}\iota b$. The damping factor is

$$\exp\,(-\tfrac{1}{3}bpt) \text{ or } \exp - \tfrac{1}{3}\left(\frac{t}{\mathrm{t}_1} + p^2 t\mathrm{t}_2\right) \qquad \text{.........(4).}$$

Since the motion has lasted for 30 years we can reasonably suppose that t_1 is at least 10 years or 3×10^8 sec. Comparing this with 14·424 (5) we see that the absorption of seismic waves can be attributed only to elastic

* *M.N.R.A.S. Geoph. Suppl.* 1, 1926, 416.

afterworking and not to plasticity. On the other hand if we take $t_2 = 0{\cdot}004$ sec. it follows that for this motion in 30 years

$$\tfrac{1}{3}p^2 t t_2 = 4 \times 10^{-9} \qquad \text{.........(5)},$$

and the effect of elastic afterworking on the variation of latitude is insignificant.

The maximum possible viscosity of the core can also be shown to imply a damping factor of the order of $\exp(-t/1{\cdot}6 \times 10^{11}$ days), which is again insignificant.

The persistence of the variation of latitude and the absorption of seismic waves therefore imply together that in the rocky shell

$$t_1 > 3 \times 10^8 \text{ sec.}; \quad t_2 = 0{\cdot}004 \text{ sec.} \qquad \text{.........(6)}$$

approximately. If the former is combined with a mean rigidity of $1{\cdot}7 \times 10^{12}$ dynes/cm.^{2} for the shell, it implies a viscosity over 5×10^{20} c.g.s.

14·427. Now in the bodily tide we have a motion whose amplitude is practically proportional to $1/\mu$, and therefore to $1 - \iota\left(pt_2 + \dfrac{1}{pt_1}\right)$, where $2\pi/p$ is now half a lunar day. Thus the bodily tide should show a lag in phase equal to

$$pt_2 + \frac{1}{pt_1} \qquad \text{.........(1)},$$

the first term in which is 6×10^{-7}, and the second under 2×10^{-5}. The sum is the 2ϵ of 14·21. Substituting in 14·21 (10) and remembering that H is now 0·6 times the height of the equilibrium tide, we find

$$\frac{N}{C} = 5 \times 10^{-26}/\text{sec.}^2 \qquad \text{.........(2)}.$$

This is an insignificant fraction of the effect of tidal friction in shallow seas; bodily tidal friction is therefore of negligible importance at present.

14·5. *Tidal Friction in the Past.* If we adopt the results of 14·32, we have

$$\frac{d\Omega}{dt} = -\frac{N + N_1}{C}$$
$$= -2{\cdot}5 \times 10^{-22}/\text{sec.}^2$$

The present value of Ω is $7{\cdot}3 \times 10^{-5}$/sec. Thus Ω changes by 10^{-5} of its amount in 3×10^{12} sec., or 10^5 years. The day has therefore probably lengthened by a second in the last 120,000 years. Thus tidal friction, historically speaking, is a slow process. On the other hand, if we consider the change since the oldest known rocks were formed, and suppose tidal friction to have operated ever since at its present rate, we find that the period of rotation $1{\cdot}6 \times 10^9$ years ago must have been only about 0·84 of our present day. The ellipticity of the earth's surface corresponding to such a rate of rotation is about $\frac{1}{210}$.

14·51. If we attempt to take the extrapolation further back still, we must have recourse to equations 14·2 (3), (4), and (5). The couple N has been seen to be proportional to $(m/c^3)\, H \sin 2\epsilon$, and if we suppose ϵ to remain constant and H to be proportional to (m/c^3), as is not unreasonable, N will be proportional to c^{-6} or ξ^{-12}. It therefore increases very rapidly as the moon's distance from the earth diminishes. We shall denote its present value by N_0. The solar tidal friction is at present a small fraction of the lunar, and cannot have changed much, since the sun's distance has hardly varied. Thus it must have been unimportant in the past in comparison with lunar tidal friction. Hence we can omit equation 14·2 (4) and drop N_1 in 14·2 (5). Then

$$\frac{Mm}{M+m} c_0^2 n_0 \frac{d\xi}{dt} = N_0 \xi^{-12} \qquad \text{.........(1)},$$

$$C\frac{d\Omega}{dt} = -N_0 \xi^{-12} \qquad \text{.........(2)}.$$

If we introduce κ as in 14·3 (6), (1) becomes

$$\kappa C \Omega_0 \frac{d\xi}{dt} = N_0 \xi^{-12} \qquad \text{.........(3)},$$

whence

$$\tfrac{1}{13}(1-\xi^{13}) = -\frac{N_0 t}{\kappa C \Omega_0} \qquad \text{.........(4)}.$$

Substituting our adopted values of N_0/C, κ, and Ω, we find for the time when ξ was 0·8, and therefore the distance of the moon 240,000 km.,

$$t = -4 \times 10^9 \text{ years} \qquad \text{.........(5)}.$$

The time taken by the moon to recede from its closest approach to the earth to 240,000 km. from it would not exceed 1/20 of this. Thus the theory of tidal friction suggests an age of the moon of the order of 4×10^9 years. This is about three times the age of the oldest rocks found from radioactivity, but evidently cannot be used to do more than suggest the order of magnitude of the age of the moon.

Equations (2) and (3) together give

$$\frac{d}{dt}(\kappa \Omega_0 \xi + \Omega) = 0 \qquad \text{.........(6)}.$$

This is a form of the equation of angular momentum of the system. Evidently if Ω was once $\frac{4}{3}\Omega_0$, $\kappa\xi$ must have been $4{\cdot}82 + 1 - \dfrac{\Omega}{\Omega_0} = 4{\cdot}49$, making ξ equal to 0·93, the distance of the moon only about 27,000 km. less than at present, and the length of the month 22 of our present days. Thus the distance and period of the moon have, on the hypothesis of uniformity, not changed by large fractions of themselves during geological time.

If ζ is the elevation of the ocean tide the potential energy per unit area is $\frac{1}{2} g \rho \zeta^2$, ρ being here the density of water; the average of this over a

period is $\frac{1}{4}g\rho H^2 \sin^4\theta$. The average kinetic energy will be of the same order. The total energy of the tide is then

$$\tfrac{1}{2}g\rho H^2 . 2\pi A^2 \int_0^\pi \sin^5\theta \, d\theta = \frac{16\pi}{15} g\rho H^2 A^2$$
$$= 4 \times 10^{23} \text{ ergs.}$$

Comparing this with 14·32 (7) we see that all the energy of the tide at any moment is dissipated in the next 3×10^4 seconds—say half a day. The dissipation is therefore so heavy that friction must have a controlling influence on the tides comparable with that of inertia. The phase lag, also, must be a moderate angle. In the past, when the moon was nearer and the tides higher, the energy would tend to increase like the square, and the rate of dissipation like the cube, of the height of the tide. But energy cannot be dissipated faster than it is supplied, and the phase lag, with the most severe dissipation, could hardly exceed 45°. It seems reasonable to suppose that ϵ in the past was greater than now, but of the same order of magnitude. The estimate in (5) is then capable of some reduction.

14·52. *The Future of the Earth-Moon System.* We saw in 14·1 that a necessary condition for the existence of tidal friction is that the earth's surface shall revolve in such a way that the moon's apparent altitude, as seen from each point of the earth's surface, is continually varying. If the earth always kept the same face turned towards the moon, the tides would settle down exactly under the moon and opposite to it, and would produce no tidal frictional couple. Thus N vanishes with $\Omega - n$. In no other circumstances can N vanish. In the primitive state of the system the length of the month was probably slightly greater than that of the day, and therefore tidal friction made the moon recede; the moon will continue to recede until $\Omega - n$ vanishes, when tidal friction will vanish again. Now when this happens Ω will be equal to n, and therefore to $n_0\xi^{-3}$. By 14·51 (6), $\kappa\xi + \dfrac{\Omega}{\Omega_0}$ must still have its present value 5·82. Hence

$$\kappa\xi + \frac{n_0}{\Omega_0}\xi^{-3} = 5{\cdot}82 \qquad \text{.........(1).}$$

The smaller root of this equation is $\xi = 1/5{\cdot}1$, corresponding to the primitive state of the system; the larger is $\xi = 1{\cdot}20$. The former makes the common period of rotation and revolution 4·8 hours; the latter makes it 47 days. Since $1{\cdot}2^{13} - 1 = 10$ nearly, the time needed to approach this latter state will be of order 5×10^{10} years.

The matter will not be closed when this state is reached, however. The solar tides will continue to operate on the earth and lengthen its period of rotation. Thus the period of rotation of the earth will come to be longer than the period of revolution of the moon. A reference to 14·1 will show that when this happens the further course of the evolution is considerably altered. A fixed point P on the earth's surface will move

round more slowly than the point where the moon is in the zenith, and therefore the high tides, while still occurring after the moon has passed the zenith or nadir, will be on the opposite sides of AB from C and D. (See Fig. 16.) Thus their effect on the moon will be to retard its revolution and make it return to the earth, and the earth's rotation will at the same time be accelerated. Thus the moon will gradually return towards the earth, the earth's rotation meanwhile being retarded by the solar tides and accelerated by the lunar ones. This process will continue till the moon is at last dragged down to within a distance of the earth so short that it will be broken up by the action of the tides raised in it by the earth*; it will then ultimately form a system like Saturn's ring, but much more massive.

14·6. *The History of the Moon's Rotation.* Let us now return to equation 14·21 (10). The height of the equilibrium tide is U/g, and its amplitude is therefore $\frac{3}{4}\frac{fmA^2}{c^3 g}$, or $\frac{3}{4}\frac{mA^4}{Mc^3}$. If then the tide has approximately its equilibrium height, we shall have

$$-\frac{d\Omega}{dt}=\frac{N}{C}=\frac{18}{5}\pi f\rho \sin 2\epsilon\left(\frac{m}{M}\right)^2\left(\frac{A}{c}\right)^6 \qquad \text{.........(1).}$$

We shall have recourse to this expression again in considering the tides in the planets.

If ω denotes the moon's angular velocity of rotation and a its radius, it appears from (1) that

$$\frac{d\omega/dt}{d\Omega/dt}=O\left\{\left(\frac{M}{m}\right)^4\left(\frac{a}{A}\right)^6\right\} \qquad \text{.........(2),}$$

provided elasticity is not so great as to affect the order of magnitude of the height of the tide, that the densities are of the same order, and the phase lags of the tides also. With

$$M/m=81,\quad a/A=\tfrac{3}{11},$$

we have

$$\frac{d\omega/dt}{d\Omega/dt}=17{,}000.$$

The hypothesis that the elasticity of the moon may be neglected is valid so long as the moon was largely fluid. The densities are actually of the same order, and the phase lags will be comparable if the departures of the angular velocities of rotation from that of revolution are comparable. Thus if initially the earth and moon rotated at the same rate, this being somewhat different from the mean motion of the system, the rotation of the moon would approach the mean motion 17,000 times as fast as that of the earth would. The moon would thus be brought to present the same face always towards the earth before the rotation of the earth had been appreciably affected by tidal friction.

* Cf. Darwin, *The Tides*, 1911, 338–345. The critical distance of about twice the radius of the earth is known as Roche's limit.

The moon would be unable to retain any atmosphere or water vapour, on account of its low gravitative power. The absence of the blanketing effect of the atmosphere would enable it to solidify somewhat earlier than the earth. Henceforth any tidal friction in the moon must have arisen from imperfection of elasticity, for there can have been no seas upon it. The thermal history of the moon has probably been very similar to that of the earth, and we should therefore expect the departure of its interior from perfect elasticity to be quantitatively comparable with that of the earth. Again, the elastic tide in the moon must be much smaller than the hydrostatic equilibrium tide so far considered, having about $\frac{1}{50}$ of the amplitude. On both grounds the value of $d\omega/d\Omega$ after the solidification of the moon must have become very much smaller; it may well, however, still be as great as 100 for equal lags.

This therefore supplies us with the required explanation of the fact that the moon always keeps the same face towards the earth. The tides raised in the moon by the earth produced such friction that they made the moon's periods of rotation and revolution equal at a very early stage in its history. It is possible that ever since then, if the recession of the moon from the earth, or any internal change in the moon, made either of these periods vary, bodily tidal friction in the solid moon would commence afresh and restore the equality.

14·61. We have, however, seen that there is no reason to suppose bodily tidal friction in the earth to be perceptible, and accordingly there is no great prior probability for a hypothesis that requires it to be continually available in the moon to adjust the moon's rotation to any changes that may have occurred. Let us consider, then, what would happen if the moon was actually free from internal friction. Take an axis OA in the ecliptic, fixed in direction. Let the longest axis of the moon, which points nearly to the earth, make an angle ϕ with OA, and let the line joining the centres of the earth and moon make an angle θ with OA. Neglect the inclination of the moon's orbit and equator to the ecliptic. Put

$$\phi = \theta + \psi \qquad \text{.........(1)},$$

so that ψ is small. Now ϕ is a Lagrangian coordinate of position of the moon. If A, B, and C be the permanent parts of the principal moments of inertia of the moon about its centre, the rate of change of angular momentum of the moon about an axis through its centre perpendicular to the ecliptic is $C\ddot{\phi}$.

The couple in the plane of the ecliptic due to the earth's attraction on the moon's permanent ellipticity of figure is $-\dfrac{3fM(B-A)}{c^3}\cos\psi\sin\psi$, where M is the mass of the earth and c its mean distance. In addition there

may be a couple due to internal tidal friction. Let us, however, examine the consequences if this is absent. The equation of motion is

$$C\ddot{\phi} = -\frac{3fM(B-A)}{c^3}\cos\psi\sin\psi \qquad \text{.........(2)},$$

or, if we use (1) and neglect ψ^2,

$$\ddot{\psi} + 3n^2\frac{B-A}{C}\psi = -\ddot{\theta} \qquad \text{.........(3)}.$$

But $\dot{\theta}$ is the rate of increase of the moon's longitude, and is therefore equal to n. Thus

$$\ddot{\theta} = \frac{dn}{dt} \qquad \text{.........(4)},$$

which is very small. $\dddot{\theta}$ is of order $\frac{1}{n}\left(\frac{dn}{dt}\right)^2$. A sufficient approximation to a solution of (3) is therefore to be got by supposing that ψ also varies very slowly, so that $\ddot{\psi}$ can be neglected. Then

$$\psi = -\frac{C}{B-A}\,\frac{dn}{3n^2 dt} \qquad \text{.........(5)}.$$

Evidently $\ddot{\psi}/\psi$ is of order $\left(\frac{1}{n}\frac{dn}{dt}\right)^2$, and therefore our assumption that it is negligible leads to self-consistent results. Now at present

$$\frac{dn}{dt} = -3n_0\frac{d\xi}{dt},$$

$$\kappa\Omega_0\frac{d\xi}{dt} + \frac{d\Omega}{dt} = 0 \text{ from 14·51 (6)},$$

and therefore

$$\frac{1}{3n^2}\frac{dn}{dt} = \frac{1}{\kappa n\Omega}\frac{d\Omega}{dt}$$

$$= -4 \times 10^{-13}.$$

Again, $(B-A)/C$ is 0·00047. Thus

$$\psi = 10^{-9} \text{ nearly}$$

$$= 2'' \times 10^{-4}.$$

Thus even in the absence of any friction in the moon at present, the recession of the moon from the earth would produce a perfectly imperceptible deviation of the moon's longest axis from the line of centres.

It may be pointed out that, small though this deviation is, it has an important dynamical effect. The moon's longest axis pointing systematically to one side of the earth causes the couple on the moon produced by the earth's attraction on the moon's equatorial protuberance to be on an average negative. It is this couple that reduces the moon's rate of rotation and makes it remain equal to the rate of revolution.

14·62. It may be thought, however, that even though the moon's permanent ellipticity keeps the particular integral of 14·61 (3) small, the complementary functions may increase considerably; in other words, the amplitude of the moon's free libration in longitude may become great. This, however, is not the case. The period of the free libration is seen from (3) to be $\frac{2\pi}{n}\left(\frac{C}{3(B-A)}\right)^{\frac{1}{2}}$, or 27 months, and clearly remains for all time proportional to the moon's period of revolution. Thus the change in the period of oscillation during a complete oscillation is only a small fraction of the period itself. In these circumstances it is known* that a first approximation to the solution, valid for all time, of an equation

$$\frac{d^2y}{dt^2} + \chi y = 0,$$

where χ is a function of t,

is
$$y = A\chi^{-\frac{1}{4}} \cos \int^t \chi^{\frac{1}{2}}\, dt + B\chi^{-\frac{1}{4}} \sin \int^t \chi^{\frac{1}{2}}\, dt,$$

where A and B are arbitrary constants. Thus the amplitude is proportional to the square root of the period. If it was small at the moon's last adjustment to the hydrostatic state, which we have seen probably took place when the period of revolution was about 6·3 of our present days, it would by now have been multiplied by $(27{\cdot}3/6{\cdot}3)^{\frac{1}{2}}$ or 2·1. It would therefore still be small.

To explain the facts that the moon keeps the same face always towards the earth, and that its free libration in longitude is imperceptible, it is therefore unnecessary for tidal friction to be still operating in its interior. It is enough that tidal friction should have been sufficient to produce these conditions before, or soon after, solidification, which is highly probable; once produced, they would be permanently maintained by the earth's attraction on the moon's equatorial protuberance.

14·7. *Tidal Friction on other Planets and Satellites.* In other systems than our own, tidal friction may be expected to operate in four ways:

1. Tides raised in the satellites by their primaries will tend to make them keep the same face towards their primaries.

2. Tides raised in the primaries by their satellites will alter the rates of rotation of the primaries.

3. Tides raised in the primaries by their satellites will alter the distances of the satellites from their primaries.

4. Solar tides will affect the rates of rotation of the planets.

These four effects may be considered separately.

14·71. We notice that in 14·6 (1) m is the mass of the tide-raising body, and M and A refer to the deformed body. Thus fm/c^3, when we are

* *Proc. Lond. Math. Soc.* **23**, 1923, 428–436.

considering tides raised in a satellite, is practically n^2, and M/A^3 is proportional to the density of the satellite. With the usual assumptions as to uniformity of physical constitution among satellites, it follows that the rate of change of velocity of rotation in a satellite is proportional to the fourth power of its mean motion. We should therefore expect that all satellites whose periods are less than that of the moon would turn the same faces permanently towards their primaries; satellites of longer periods may not yet have reached this state. All satellites whose rotation periods are known do actually keep the same faces towards their primaries; they include the great satellites of Jupiter, and also Iapetus. The period of the latter is nearly three months.

14·72. It can be definitely asserted that no satellite other than the moon has produced a considerable effect on the rotation of its primary. For the effect of tidal friction is to transfer angular momentum from the rotation of the primary to the revolution of the satellite, and if the rotation of any planet had been much affected in this way the angular momentum of the satellite's revolution would be comparable with that of the planet's rotation. This is true of no satellite except the moon; the orbital angular momenta of all other satellites are insignificant in comparison with the rotational angular momenta of their primaries.

This fact is consistent with the tidal lags on the great planets being comparable with that on the earth. For if a satellite has radius a and the same density as its primary,

$$\frac{m}{M}\left(\frac{A}{a}\right)^3 = 1,$$

and therefore

$$\left(\frac{m}{M}\right)^2\left(\frac{A}{c}\right)^6 = \left(\frac{a}{c}\right)^6,$$

and thus the rate of change of angular velocity of rotation should be proportional to the product of sin 2ϵ into the sixth power of the apparent diameter of the satellite as seen from its primary. No other satellite subtends at the centre of its primary a greater angle than the moon, though the moon's apparent size is approached by Phobos and J I. It is therefore possible that the satellites have not affected the rotations of their primaries considerably, even if the primaries show as great tidal lags as the earth.

14·73. The argument is readily extended to the effect of the tides raised by the sun. The sun subtends at Jupiter an angle of only 6′, which is less than that subtended by J I, whose density is very similar. The lags of the tides in Jupiter raised by the sun and J I cannot be very different. Thus the effect of solar tidal friction on Jupiter must be less than that of J I, which we already know to be insignificant. The effects on Saturn, Uranus, and Neptune must be still smaller.

Mars is probably more nearly comparable with the earth, and tidal dissipation on it may occur in shallow seas. The apparent diameter of the sun as seen from it is 21′, and that of Phobos, inferred from the brightness, is about 20′. Phobos, however, probably has a greater density than the sun, and thus the solar tides are probably less important to Mars than those raised by Phobos. They therefore can hardly have affected the rotation of the planet much.

Venus and Mercury are in a different position. Supposing the tidal lags in these planets to be equal to that in the earth, we see that the rate of change of Venus's speed of rotation must be $(0{\cdot}72)^{-6}$, or 7·2 times the effect of the sun on the earth. The corresponding ratio for Mercury is about 300. Venus has an atmosphere and may have shallow seas; Mercury, on the other hand, has none, and friction in it, if any, must be bodily. It appears probable, then, that if Venus were now rotating in the same time as the earth, its rate of angular retardation would be rather more than twice that of the earth on account of the sun and moon together; but this rate is not nearly so fast as the former rate of angular retardation of the earth when the moon was nearer. If then Venus once had a short period of rotation, of a few hours only, it is probable that it would not yet have been made to rotate in so long a period as the earth; if, however, it had a period such as the earth has now, it would have been very much lengthened by solar tidal friction. It is now known from spectroscopic observations* that the rotation of Venus cannot be nearly so fast as that of the earth, so that we may infer that its original period of rotation was probably not very fast.

Mercury has probably had a history similar to that of the moon. Its periods of rotation and revolution were made equal by solar tidal friction, probably before it was thoroughly solid, and this condition has been maintained ever since by bodily friction; the distance of this planet from the sun cannot have changed much, but an ellipsoidal inequality of the planet's figure may have been produced in solidification in the way suggested for the moon, and have been afterwards maintained by the strength of the material. This may have since kept the same face of Mercury turned towards the sun, in the manner suggested for the moon in 14·61.

14·74. The rate of increase of a satellite's distance through tidal friction is given by 14·2 (3)

$$\frac{Mm}{M+m}c^2 n\frac{d\xi}{dt} = N \qquad \text{.........(1).}$$

From 14·6 (1)

$$\frac{N}{C} = \tfrac{18}{5}\pi f\rho \sin 2\epsilon \left(\frac{m}{M}\right)^2 \left(\frac{A}{c}\right)^6 \qquad \text{.........(2),}$$

and we know that for the earth $C = \frac{1}{3}MA^2$ ....(3).

For other planets the numerical coefficient is not very different.

* Russell, Dugan, and Stewart, *Astronomy*, 1926, 317.

On eliminating N and C we find

$$\frac{d\xi}{dt} = \tfrac{6}{5}\pi f\rho \sin 2\epsilon \frac{(M+m)\,m}{M^2 n}\left(\frac{A}{c}\right)^8$$

$$= \tfrac{6}{5}\pi f\rho \sin 2\epsilon \frac{m}{Mn}\left(\frac{A}{c}\right)^8 \qquad \text{.........(4),}$$

if we treat $(M + m)/M$ as equal to unity.

For J I, m/M is 1/20,000, as against 1/80 for the moon; n has about 20 times its value for the moon; A/c has about 10 times its value for the earth and moon. Thus $d\xi/dt$ has about 2×10^4 times the value corresponding to the moon at present. If then the constitution of Jupiter resembled that of the earth, the mean distance of J I would be doubled in about 6×10^5 years.

For Titan, again, $d\xi/dt$ should be about 60 times what it is for the moon: and for the satellite of Neptune it should be about 2000 times as great.

It appears then that unless the great planets approximate very closely to perfect elasticity, giving much smaller tidal lags than the average in our ocean, the evolution of their nearer large satellites must have been dominated by tidal friction. This will not be true of the more remote satellites, since the factor $(A/c)^8$ diminishes very rapidly as c increases.

Phobos presents a further difficulty. It is readily seen that, on the same physical hypotheses, $d\xi/dt$ for it should be 8000 times as great as for the moon. This result is even more embarrassing than that for J I; for Phobos revolves more rapidly than Mars rotates, and therefore tidal friction, in accordance with the argument of 14·52, will make Phobos approach the planet instead of receding from it. Thus Phobos would have been precipitated on the surface of the planet ere now if it was as old as the moon*.

It is possible, however, that our estimate will have to be much reduced, for two reasons. Mars may have little fluid on its surface, and therefore no dissipation in its shallow seas. If so, tidal friction within it must be bodily. A reduction of $d\xi/dt$ to a small fraction of the value just found may therefore have to be made. A further reduction will be needed on account of the small size of Mars. In a small planet the elastic tide is a small fraction of the hydrostatic tide; thus in Mars the height of the tide may well be reduced to $\frac{1}{20}$ of its hydrostatic value by rigidity. For these reasons it is quite possible that the evolution of the orbit of Phobos may proceed no faster than that of the moon.

We notice, however, that the arguments just used for Mars are applicable with greater force to Mercury. Thus if the solar tides are supposed to have made the latter planet keep the same face turned to the sun, we seem to be driven to adopt the theory of 14·61 for Mercury as well as the moon.

* Cf. James Nolan, *Nature*, **34**, 1886, 287.

14·8. *Summary*. Friction, either in oceanic or in bodily tides, must produce a continual diminution in the rate of rotation of the earth, and continual increases in the mean distances of the moon and sun from the earth. These changes together make the moon and sun appear to have, relative to the stars, slow secular accelerations which can be found by comparison of modern observations with ancient ones of eclipses and occultations. The secular acceleration of the sun found in this way is a somewhat larger fraction of that of the moon than would theoretically have been expected. The rate of dissipation of energy required to explain these accelerations is about $1{\cdot}4 \times 10^{19}$ ergs per second on an average.

It is found that tidal friction in the open ocean cannot account for more than about a thousandth of this, but that the greater part of it can be explained quantitatively by tidal friction in shallow seas. There is no reason to believe these seas unable to account for the whole of it, and therefore tidal friction in the body of the earth may be inappreciable. It appears, indeed, that if a considerable fraction of the secular accelerations were due to elastic afterworking, earthquake waves could not be transmitted; and that if it were due to plasticity in a homogeneous earth, the 14-monthly variation of latitude could not exist.

By extrapolation it is found that the period of the earth's rotation may have changed by 4 hours or more during geological time. Tidal friction is capable of explaining how the moon came to recede from close proximity to the earth to its present distance; the time required does not appear likely to be prohibitive. The moon will ultimately recede till its period of revolution and the period of the earth's rotation are both equal to about 47 of our present days. When this takes place the moon will gradually approach the earth again, ultimately passing within Roche's limit and being broken up. (It may be remarked that on the resonance theory, on account of the great extension of the earth at the time, the moon was outside Roche's limit when it was first formed.)

Tidal friction readily accounts for the fact that the moon always keeps the same face towards the earth; it is sufficient that this condition should have been brought about before the moon solidified, for afterwards it could have been maintained by purely gravitational causes. In the same way, Mercury has probably been made to keep the same face towards the sun. This result is probably applicable to all satellites whose periods are less than that of the moon. No planet except the earth has had its rotation much affected by tides raised by a satellite. It is difficult to make any inference about Venus, except that its rotation period has not been lengthened to such an extent as that of the earth. The orbits of some satellites may have been appreciably affected by tidal friction, notably J I and Phobos, but further inferences cannot be made without more knowledge about the physical conditions on their primaries.

CHAPTER XV

The Origin of the Earth's Surface Features

"You couldn't deny that, even if you tried with both hands."
"I don't deny things with my *hands*," Alice objected.
"Nobody said you did," said the Red Queen. "I said you couldn't if you tried."
"She's in that state of mind," said the White Queen, "that she wants to deny *something*,—only she doesn't know what to deny!"
"A nasty, vicious temper," the Red Queen remarked.

LEWIS CARROLL, *Through the Looking-Glass.*

15·1. The conspicuous superficial phenomena that require physical explanations are the difference between continents and ocean basins, the formation of mountains, and the various types of igneous activity. Mountain formation being considered first, the folding and thrusting shown in existing ranges imply a definite shortening of the crust; this in turn implies a reduction of the earth's surface and therefore of its volume. The existence of folded mountains implies a contraction of the earth's interior. It is worth while to emphasize this at the start, since some recent writers have attacked the contraction theory in general, apart from any special view as to the cause of the contraction. Even if the contraction theory was given up, it would still be necessary to find a theory to account for the contraction.

Of the many causes of mountain ranges that have been proposed, only two have been shown adequate to account for any appreciable fraction of the crumpling that has occurred, and many are even qualitatively as well as quantitatively unsatisfactory. The most effective seem to be thermal contraction and changes in the rotation of the earth.

15·2. *Thermal Contraction.* It has been seen in Chapter VIII that different parts of the interior of the earth have cooled since solidification by different amounts. In cooling they must have contracted in volume in different ratios, and in this way a state of stress must have been set up in the crust. The mathematical discussion of the character of the deformations produced was due originally to Dr C. Davison* and to Sir G. H. Darwin†. Consider the earth at some instant during its cooling, and consider the effect of the cooling that takes place during a further interval. Throughout the region from the centre of the earth to within about 700 km. of the surface, no appreciable change of temperature takes place, and therefore no change of volume. Between this level and the

* *Phil. Trans.* A **178**, 1887, 231–242.

† *Ibid.* A **178**, 1887, 242–249; or *Sci. Papers*, **4**, 354–361. The contraction theory qualitatively goes back to Newton; cf. Brewster, *Memoirs of Sir Isaac Newton*, **2**, Appendix 4.

layer where cooling is most rapid, each layer cools more than the layer below it, and would therefore contract more if it were not obstructed by the matter below. The latter fixes the inner radius of this region, and therefore the requisite reduction in volume can be achieved only by reducing the outer radius. Thus the adjustment requires a thinning of this region without a corresponding reduction in its inner radius. Since this region is necessarily in the region of low strength, the matter in it will adjust itself to the stresses involved, and assume a hydrostatic state. On the other hand, the outer surface of the earth undergoes no further cooling and contraction, and is therefore too large to fit the contracted region just considered. It will therefore be under a horizontal crushing stress.

Since the region below the layer of greatest cooling becomes too small to fit the interior, while the outer surface becomes too large, there must be an intermediate layer where the contraction is just enough to enable it to continue to fit the interior. This layer is called the 'level of no strain.'

Below the level of no strain, the rocks at any time would, if they simply underwent a contraction in all directions in accordance with their cooling, be too small to fit the interior, and thus in order to fit it they are stretched out horizontally. Yield under horizontal extension may occur in two ways. If the rock is capable of flow, it merely spreads out horizontally and becomes thinner vertically. It may, however, fracture vertically, forming long fissures, which will descend to levels where flow is possible. In such a case, hydrostatic pressure will force the matter capable of flow up into the crack, and its further cooling will lead to its solidification or crystallization within the crack. In either case, the total volume of the matter within the level of no strain is the same, for the vertical crack cannot extend above this level, and therefore the motion of weak material into the crack only redistributes matter below the level of no strain without altering its quantity.

Above the level of no strain the rocks have probably considerable strength. Hence when exposed to horizontal stress they would not flow, but would fracture and bend up irregularly, starting at the weakest spot. Thus a system of folds would be produced in the upper rocks. These, according to the thermal contraction theory of mountain building, are the initial stages of mountain ranges.

Since there is no deformation initially, and cooling starts from the surface, the level of no strain must start from the surface and gradually descend. By the time it has reached any particular layer of material within the crust, that layer will be solid, and therefore will yield to subsequent deformation by folding. This folding will start at once, for the filling up of any fissures that may have been opened will have blocked them up with solid material; thus when compression is applied the rocks will not be free to move, and yield can take place only by crumpling.

Accordingly the crumpling of any layer required to make it continue to fit the interior must be calculated from the time when the level of no strain passed that layer.

15·21. *The Amount of Compression available.* Let us now apply the above considerations in a quantitative discussion of the amount of crumpling that must have occurred on the earth, to enable the outside to continue to fit the interior. Consider a shell of internal radius r and thickness dr. Let its coefficient of linear expansion be n, where n may be variable, and let its initial density be ρ. Let the *rise* of temperature of the shell be v; this will of course be negative. Then v is a function of r alone. The density of the shell, if we ignore the small change due to compressibility, becomes $\rho\,(1 - 3nv)$. Let the radius of the shell of radius r become $r\,(1 + \alpha)$. Then the external radius becomes

$$r\,(1+\alpha) + dr\left\{1 + \frac{d}{dr}(r\alpha)\right\} \qquad \text{.........(1).}$$

Hence the mass of the shell after the change of temperature is

$$4\pi r^2\,(1+\alpha)^2 dr\left\{1 + \frac{d}{dr}(r\alpha)\right\}\rho\,\{1 - 3nv\} = 4\pi\rho r^2 dr\left\{1 + 2\alpha + \frac{d}{dr}(r\alpha) - 3nv\right\} \qquad \text{.........(2),}$$

neglecting squares and products of α and v. But the mass is unaltered. Hence we have the equation of continuity

$$2\alpha + \frac{d}{dr}(r\alpha) - 3nv = 0 \qquad \text{.........(3).}$$

Since v is supposed known throughout the earth, this is a differential equation to determine α.

Now if a shell simply expanded without stretching, its radius would increase by rnv instead of $r\alpha$, so that the stretching required to make it continue to fit the interior is $r\,(\alpha - nv)$. Let us denote $\alpha - nv$ by k. Then substituting for α in (3) we have

$$\frac{d}{dr}(kr^3) = -\,r^3\frac{d}{dr}(nv) \qquad \text{.........(4).}$$

At the centre there will be simple expansion without stretching, so that k will vanish with r. Hence

$$k = -\frac{1}{r^3}\int_0^r r^3\frac{d}{dr}(nv)\,dr \qquad \text{.........(5).}$$

Let R be the radius of the earth. Then we can write

$$r^3 k = -\,[r^3 nv]_0^r + \int_0^r 3r^2 nv\,dr \qquad \text{.........(6),}$$

$$k = -\,nv + \frac{1}{r^3}\int_0^r 3r^2 nv\,dr \qquad \text{.........(7),}$$

since the integrated part vanishes at the lower limit. Now the change of temperature is appreciable only in a depth small in comparison with the

radius of the earth. Hence in the region where k is appreciable r can to a first approximation be put equal to R. Then

$$k = -nv + \frac{3}{R}\int_0^r nvdr \qquad \text{.........(8).}$$

If we call the depth x, we have

$$r = R - x \qquad \text{.........(9),}$$

and then, since the temperature change is practically that in a solid of infinite depth with a plane face, as was seen in Chapter VIII, we have

$$k = -nv + \frac{3}{R}\int_x^\infty nvdx \qquad \text{.........(10).}$$

Now consider the changes that take place in a short interval of time dt. If the integral stretching in a particular layer since solidification be K, we have

$$k = \frac{\partial K}{\partial t}dt; \quad v = \frac{\partial V}{\partial t}dt \qquad \text{.........(11),}$$

where V is the temperature. Then (10) becomes

$$\frac{\partial K}{\partial t} = -n\frac{\partial V}{\partial t} + \frac{3}{R}\int_x^\infty n\frac{\partial V}{\partial t}dx \qquad \text{.........(12).}$$

The stretching at the surface is to be found by putting x zero in this and integrating with regard to the time. The level of no strain is determined by the fact that $\partial K/\partial t$ vanishes there.

15·22. Unfortunately the experimental determination of the coefficients of expansion of rock materials is difficult. They usually disintegrate when heated to some temperature near 800° C., and show an apparent expansion due only to the formation of cavities. This change is not reversed on cooling. If originally vitreous they may partially crystallize, with contraction in volume, when heated to a temperature where molecular mobility is considerable. In former works of mine Fizeau's determinations of the coefficients of expansion were used; according to these

$$n = \epsilon + \epsilon' V,$$

where $\epsilon = 7 \times 10^{-6}/1° \text{ C.}; \; \epsilon' = 2{\cdot}4 \times 10^{-8}/(1° \text{ C.})^2$,

giving a linear expansion of 1·9 per cent. between 0° and 1000° C. N. E. Wheeler* gives 2·73 per cent. for granite and 1·54 per cent. for olivine dolerite. Day, Sosman and Hostetter, in a critical investigation†, find a complicated behaviour at high temperatures, with volume expansions of the order of 10 per cent. between 900° and 1250° C. But the results can give only the order of magnitude of the actual contraction within the earth, where the pressure presumably prevents the formation of cavities and much of the matter may be vitreous. The results of 8·4 showed that an average cooling of about 500° is to be expected through depths down to 400 km. This,

* *Trans. Roy. Soc. Canada*, **4**, 1910, 19–44. † *Amer. J. Sci.* **37**, 1914, 1–39.

with a volume contraction of 5 per cent., would imply a reduction of 20 km. in radius and therefore of 130 km. in circumference. To this should be added allowances for contraction in crystallization and for loss of volume due to extrusion of volatile constituents, especially water. The cooling below the oceans is probably greater than below the continents. The uncertainties in all the determinations, however, are such that we can only say that the compression of the crust is of the order of 200 km.; it may easily be half or twice this. The depth of the level of no strain is about 100 km.

15·23. Direct measurement of the folding has been carried out in several of the great mountain chains of the globe. Up to a few years ago estimates usually quoted were 40 to 50 miles in the Appalachians, 25 miles in the Rocky Mountains of British Columbia, 10 miles in the Coast Range of California, and 74 miles (118 km.) in the Alps*. The tendency of later work has been to discover new thrust planes and to increase all these estimates. Thus Heim† gives 200–300 km. for the compression involved in the Tertiary folding of the Alps; A. Keith gives 320 km. for the Appalachians‡, and G. P. Mansfield mentions one thrust with a displacement of 35 miles (56 km.) in the Rockies§. If the shortening of the crust is 200 km., the reduction in the area of the earth's surface is 5×10^{16} cm.^{2} Using the earlier estimates of compression in the surveyed ranges, and inferring those for other ranges from considerations of similarity, I found that the areal compression needed to account for all the mountains of the globe was about 2×10^{16} cm.^{2}||, subject to some increase to allow for old ranges already denuded away. The available compression would therefore appear ample, and perhaps indeed embarrassingly superfluous, if the geological estimates of the observed compression had not recently been so much increased.

15·24. There is, however, another line of evidence that seems to indicate that part of the observed folding and thrusting must be due to something besides crustal shortening. Suppose the normal structure of a continent to be represented by 1 km. of sediments of density 2·4, 10 km. of granite of density 2·6, 20 km. of tachylyte of density 2·9, the whole resting on dunite of density 3·3. Imagine compression enough to double the thickness of all the upper layers, giving an increase of 31 km. The extra weight would displace 26·2 km. of the lowest layer, leaving an elevation of 4·8 km. The height of the summit of Mont Blanc above the North Italian plain is 4·5 km. It seems clear that a shortening in the ratio 2 : 1 would give a

* Pirsson and Schuchert, *Textbook of Geology*, 1915, 361; repeated (except for the Alps) in the second edition, 1920.

† Albert Heim, *Geologie der Schweiz*, **2**, 1921, 50.

‡ *Bull. Geol. Soc. Amer.* **34**, 1923, 335. § *Ibid.* 268.

|| *Phil. Mag.* **32**, 1916, 575–591.

general elevation higher than the maximum height of the Alps, and far more than their average height. Yet Heim's estimate involves a shortening in a ratio of 3 or 4 to 1, and even higher values are given by other writers.

Such a result makes it necessary to reexamine our premises. The present relief in mountainous regions is due mainly to the carving of deep valleys, at the expense, in the first instance, of the sedimentary layer. The mean thickness of the light layers is therefore reduced, and isostatic balance is restored by the inflow of new heavy matter. The loss of a kilometre from the top implies the addition of 0·73 km. below, so that the mean level is lowered by 0·27 km. But the denudation is mainly at the expense of the mountain slopes and valleys; if the mountain tops had been denuded by the average amount for the region they would never have become mountain tops. The inflow by itself would raise the mountain tops by 0·73 of the average depth denuded, and if the summits themselves are denuded less than this the effect of denudation and compensation together will be to raise the summits higher. Presumably the ratio is much less than 0·73, and we may reasonably infer that the height of the mountains above sea level is now greater than when they were formed, by an amount of the order of half the mean depth removed, or, say, a quarter of the elevation of the mountain tops above the neighbouring valleys*.

In the Alps and Rockies the height of most of the highest peaks is about 4 km., and that of the valleys about 1 km. In the Himalayas the corresponding values are about 7 km. and 2 km. Hence we may suppose that the primitive height of the Alps and the Rockies was about 3·2 km. and that of the Himalayas 5·8 km. The shortening of the crust is then in the ratio 1·66 : 1 for the Alps and Rockies and 2·2 : 1 for the Himalayas. Allowance for denudation therefore makes the discrepancy worse.

On Holmes's view that the intermediate layer is diorite of density 2·8 and the lower eclogite of density 3·4, doubling the thickness of the crust would give a residual elevation of 6·2 km. instead of 4·8 km. This hypothesis therefore also increases the discrepancy.

Nor is any help to be obtained from the hypothesis of S. Mohorovičić that the granitic layer is 60 km. thick. Doubling its thickness would then give an elevation of no less than 12 km., so that the crustal shortening needed to form the Alps would be only 0·27 of the present width, as against Heim's factor of 2 or 3. Replacing the lowest 20 km. of the granite by basalt reduces the discrepancy only a little. But in any case this hypothesis is directly contradicted by the geodetic results. The thickness in a mountain region would have to be more than the normal 55–60 km., while compensation in the Alps appears to be at a depth of only 41 km. Even without any allowance for regionality of compensation Mohorovičić's depth is too great to agree with the facts.

* *Gerlands Beiträge*, **18**, 1927, 1–29; the point has also been noticed by Nansen, *Avhand. Norsk. Videns-Akad.*, Oslo, 1927, **92**, No. 12.

15·241. The hypotheses of other writers concerning the structure of the upper layers are therefore even more discordant than mine with the estimated crustal shortening in the great mountain ranges. The only apparent hope of improvement would lie in a substantial reduction in the adopted density of the lower layer, to, say, 3·0 instead of 3·3 gm./cm.^{3} The objection to this is that no known rock of this density would give the observed velocities of P and S waves; and as all the minerals contained in common rocks near this density have been tested it seems wildly unlikely that any other possible rock would do so.

15·25. It must be recalled that the ranges in question include those where isostasy is best established by observation, and the failure of recent estimates of compression in them to agree with those indicated by isostasy can have only one explanation: that the compression measured in the field includes something that is not crustal shortening, and is therefore unsuited for comparison with the shortening indicated by the thermal contraction theory.

15·26. On the other hand we may now use the estimates derived from isostasy directly. The present width of the Alps near the Jungfrau and of the Rockies near Pike's Peak may be taken as 100 km., and that of the Himalayas near Mount Everest as 160 km. The compression involved in their formation is therefore about 70 km. for the Alps and Rockies, and 190 km. for the Himalayas. But alternatively the Himalayas may have been formed by a smaller compression acting on a greater thickness of light rocks, possibly sedimentary.

There have been several epochs of mountain formation during the earth's history comparable in magnitude with the Tertiary one that produced the mountains just considered. On the whole it seems that the agreement between the actual and predicted amounts of compression is as good as could be expected.

15·27. *The Pacific Mountains*. The estimate of the available compression has been based on the curve on p. 154, which is adapted to continental conditions. Under the oceans there must be a range of depth where the cooling has been notably more than below the continents*. We cannot say yet whether this range is to be measured in tens or hundreds of kilometres, but the essential point here is its existence. In addition, it appears that basic rocks are, on the whole, stronger than acidic ones; thus basalt has, under ordinary conditions, a crushing strength of $1{\cdot}2 \times 10^9$ dynes/cm.^{2}, as against $0{\cdot}8 \times 10^9$ dynes/cm.^{2} for granite†. On account both of the greater intrinsic strength and of the lower temperature, the rocks below the oceans must be stronger than those below the continents. Now where the compressed ocean floor abuts on a compressed continent, the weaker will be the first to give way; the continent margin

* Cf. p. 159. † Landolt and Börnstein, *Physikalische Tabelle*.

will be forced inwards and its rocks piled up over those further inland, forming ranges of mountains parallel to the coast. These correspond closely to the mountains of the Pacific Coast of America. They are evidently, if this theory is correct, formed by the relief of sub-oceanic and not sub-continental compression, so that the compression produced has been derived from the oceanic rocks.

15·28. *Epochs of Diastrophism*. The formation of mountains has not taken place at all periods in the history of the earth; it is known that there have been long intervals of quiescence. These appear as a natural consequence of the present theory. In granitic rocks the modulus of rigidity is about 3×10^{11} dynes/cm.^{2}, and Young's modulus is about 8×10^{11} dynes/cm.^{2} The breaking stress being taken as 8×10^8 dynes/cm.^{2}, we see that if Hooke's law held right up to the breaking-point, the extension would be -10^{-3}. But it has been seen that the linear compression to be expected on the contraction theory is about 5×10^{-3}. Hence the compression at any place has had time to reach the breaking stress and undergo complete relief about five times.

What must happen on the thermal contraction theory of mountain formation (or indeed on any other theory that takes account of the properties of solids) is that the stresses increase continuously until the strength of the rocks is reached, when flow commences. Until this stage no flow and no mountain building occur, and we have an interval of quiescence. When the stress-differences reach the strength, complete fracture takes place in surface rocks, and the strength is reduced to zero. Thus crumpling continues until the stresses are almost completely relieved. This corresponds to an epoch of mountain formation. Then the fractures become sealed up afresh, and further internal cooling recommences the process. By what has just been said about the strength of rocks, there may have been about five such epochs since the solidification of the earth. It is of interest that this is of about the order of the number of the great eras of mountain-building that are geologically known to have occurred.

15·29. It has been seen on page 279 that the thinning of the crust below the layer of no strain is an essential preliminary to mountain formation. The depth of this layer is of the order of 100 km.; that of the asthenosphere is of order 40 km. (Though the thickness of the lithosphere is of the same order of magnitude as the combined thickness of the two upper layers of seismology, there is not yet any definite reason for identifying them.) Thus the initial thinning would take place under very small stress-differences. The outer layers are pulled down on to the thinned region by gravity until the stresses produced are greater than they can support, and fracture takes place. The piling up of the crumpled rocks necessarily increases the weight of rock per unit area locally, and hence there is a local increase in the pressure on the asthenosphere, which produces enough outflow to reduce the stress-differences there to the strength. But the horizontal

movements involved in the crumpling themselves involve motion of the lithosphere over the asthenosphere and are resisted by the viscosity of the latter until it is pushed out of the way. It appears therefore that the formation and compensation of mountains are parts of the same process and proceed simultaneously.

15·3. Some of the alleged objections to the contraction theory in general and the thermal contraction theory in particular have been met in the course of the above account. Some have been answered several times already, but appear to be capable of indefinite repetition however often they are answered.

15·31. It is often said that continuous contraction of the interior would not give mountain formation on a world-wide scale, but only a vast number of very minute puckers. This really involves two questions: whether the formation of a few great mountain girdles is capable of reducing the compressive stresses below the strength of surface rocks, and whether any circumstance in the nature of the process indicates that relief would actually form such girdles. With regard to the first, we may notice that if the earth was flat a uniform compressive stress over a square continent could be relieved completely by two folds, one parallel to each pair of sides. Any tendency to general puckering could then arise only as a secondary effect due to the curvature of the earth. Now imagine a circular continent subtending an angle 2α at the centre of the earth; the radius of the earth being R, the diameter of the continent is $2R\alpha$, and its perimeter $2\pi R \sin \alpha$. When α is small any change of R and α affects the diameter and perimeter of the continent in the same ratio. For continents of finite size the ratios may differ by quantities of the order of $(1 - \cos \alpha)\, \delta\alpha$. For an uncompressed continent $\delta\alpha$ would be $\alpha \delta R/R$, and the actual value will be less. The difference between the extensions along the radii and perimeter of a continent during distortion is therefore at most of the order of $1 - \cos \alpha$ of either separately. For a continent 4000 km. wide this is only $\frac{1}{20}$. Thus when stresses even in a large continent reached the strength they could be relieved by the formation of two intersecting ranges, each involving a shortening of the crust in the same ratio, and leaving only a small residual stress unrelieved till the next general upheaval. There seems to be no reason why the ranges should be more closely spaced than they actually are*.

When failure first takes place in a solid it produces a local departure from uniformity of structure, which constitutes a flaw. Near any such flaw the local stresses considerably exceed the general ones, and consequently, as is well known, any fracture or fold tends to extend. Thus when the stresses first reached the strength of the crust they would produce deformations that would extend widely, forming long ridges, until the stresses were sufficiently relieved to obviate the need for further shortening.

* It is really for similar reasons that it is possible to make a terrestrial globe by fitting to a sphere sectors printed on plane paper.

The formation of a few great ranges is then to be attributed simply to the fact that it is easier to extend an old break than to start a new one.

15·32. The objection that continuous cooling would give continuous adjustment and not long quiet intervals separated by short and great upheavals has been answered already in 15·28. The actual alternation is precisely what would be expected when the finite strength of ordinary solids is remembered. Indeed the fact that a compression of the order of 1 part in 1000 is required to produce set in granite implies directly that the effect of a general crustal compression in a sphere 4×10^4 km. in circumference would be to shorten the crust by something of the order of 40 km. whenever yield took place. This is comparable with the amount of compression shown in typical great ranges.

15·33. It has also been said that a thin solid crust could not transmit thrusts over long distances, but would be buckled into ridges of small horizontal extent. The question is one of what is called elastic instability, and has numerous applications in engineering. A long bar or beam under longitudinal thrust, if not prevented from bending sideways, will do so under a thrust much less than is needed to produce fracture directly. The condition for this to happen is that the work done by the thrust on the ends, when the ends are displaced inwards, shall exceed the elastic energy of bending. But in the case of the earth's crust the work done against gravity and the pressure of the asthenosphere below must be allowed for, since the tendency of the asthenosphere to return to the spherical form under gravity and hydrostatic forces alone implies a definite opposition to bending of the crust, and the theory of the buckling of a free beam does not apply directly.

We may neglect the curvature of the earth and treat the problem of a beam under two-dimensional distortion in a vertical plane. The horizontal distance from the end is x, the vertical displacement y (assumed small compared with the length). The length is

$$\int \left\{1 + \left(\frac{dy}{dx}\right)^2\right\}^{\frac{1}{2}} dx$$

taken between the ends. If then the strain is such as not to alter the length when the beam is bent, the ends must approach by a distance

$$\tfrac{1}{2}\int \left(\frac{dy}{dx}\right)^2 dx.$$

If Q is the thrust on the ends of the beam, it therefore does work

$$\tfrac{1}{2}Q \int \left(\frac{dy}{dx}\right)^2 dx \qquad \text{.........(1).}$$

The elastic energy involved in bending is*

$$\tfrac{1}{24} Ed^3 \left(\frac{d^2y}{dx^2}\right)^2 \text{ per unit length,}$$

* Cf. Lamb, *Statics*, 1921, 313.

where E is Young's modulus for the crust and d its depth. When the crust is elevated a distance y a force $g\rho y$ per unit area is needed to support the weight of the extra column of the underlying material of density ρ. The gravitational energy is therefore $\frac{1}{2}g\rho y^2$ per unit area. Our condition for instability is therefore

$$\int\left[\tfrac{1}{2}Q\left(\frac{dy}{dx}\right)^2 - \tfrac{1}{24}Ed^3\left(\frac{d^2y}{dx^2}\right)^2 - \tfrac{1}{2}g\rho y^2\right]dx > 0 \quad \text{.........(2).}$$

It follows at once that if the disturbance is of very short wave-length the elastic term will be larger than the work done by the thrust. If the wave-length is very long the gravity term will be the greatest. In general if the wave-length is $2\pi/\kappa$, and the region has a horizontal extent of many wave-lengths, so that end corrections may be omitted*, the condition is that

$$Q\kappa^2 > \tfrac{1}{12}Ed^3\kappa^4 + g\rho \quad \text{.........(3),}$$

which is satisfied by some real value of κ provided

$$Q^2 > \tfrac{1}{3}Ed^3g\rho \quad \text{.........(4).}$$

Now Q cannot exceed the thrust involved if a stress equal to the crushing stress of granite, say 10^9 dynes/cm.^{2}, exists through a depth d; that is, $Q \leqslant 10^9 d$. E can be taken as the Young's modulus for granite, about $7{\cdot}5 \times 10^{11}$ dynes/cm.^{2}, and ρ is the density of the matter in the asthenosphere, about 3·3 gm/cm.^{3} Substituting these values we find that we cannot produce elastic instability before fracture unless

$$d < 1{\cdot}2 \times 10^3 \text{ cm.}$$

If the crust transmitting the thrust is more than 12 metres thick, the combined effect of elasticity and gravity will be to keep it thoroughly stable for any thrust capable of being borne by the material at all. It seems to be a matter of minor importance, in view of the order of magnitude of this result, whether the 'crust' for this purpose is taken to mean the lithosphere, the two upper layers of seismology, or the region above the level of no strain. In any problem of mountain formation the thrust can transmit the stresses perfectly for any distance, and failure takes place where the stress-difference first reaches the strength of the rocks.

15·34. A further objection that has been urged against the contraction theory in general is that it supposes the interior to contract away from the outside, leaving the latter unsupported, whereas it is easy to show that an unsupported spherical shell would be put by gravity in a state of stress such as no known material could stand. It is difficult to see why this should be urged as an objection to the theory, seeing that the fact that the outer crust has been subjected to stresses that it was unable to stand is the very thing we have to explain; and it would be interesting to know what the objectors think *would* happen if the interior of the earth

* The complete problem, with proper allowance for the end conditions, is discussed by S. Goldstein, *Proc. Camb. Phil. Soc.* 23, 1926, 120–129.

underwent a radial contraction of 20 km. But actually the idea that the outside is left unsupported is not a part of the theory, as will be seen from the following investigation of the stresses in a contracting globe before set has occurred.

If the difference between continents and oceans be ignored, and the earth be considered simply as cooling from the outer surface, let us consider the elastic strain due to the change of temperature. The earth in these conditions will remain perfectly symmetrical, and if the centre be taken as the origin of coordinates, we have, with the notation of Chapter IX,

$$u = qx/r, \quad v = qy/r, \quad w = qz/r \qquad \text{.........(1)},$$

where q is the radial displacement and r the distance from the centre.

The radial force acting on unit mass is g, and is of course in general a function of r. Thus

$$X_0 = -\, gx/r, \text{ etc.} \qquad \text{.........(2)},$$

and

$$uX_0 + vY_0 + wZ_0 = -\, gq \qquad \text{.........(3)}.$$

The force per unit mass at (x, y, z) after the displacement is

$$\left(X_0 - q\frac{dX_0}{dr}\right)\left(\frac{r-q}{r}\right)^2, \text{ etc.},$$

so that

$$X_1 = \frac{2gxq}{r^2} + \frac{qx}{r}\frac{dg}{dr} \qquad \text{.........(4)}.$$

Also

$$\delta = \frac{1}{r^2}\frac{d}{dr}(r^2 q) \qquad \text{.........(5)},$$

$$\rho_1 = -\,\rho_0\delta - q\frac{d\rho_0}{dr} \qquad \text{.........(6)}.$$

Substituting in 9·1 (20), and remembering that λ and μ are functions of r alone, we find that all three equations are satisfied if

$$\frac{d}{dr}\{(\lambda + 2\mu)\,\delta\} - \frac{4q}{r}\frac{d\mu}{dr} - \frac{d\gamma}{dr} + \frac{4q}{r}g\rho_0 = 0 \qquad \text{.........(7)}.$$

If cooling has extended to a depth of order H, we see that in the portion of the earth affected by the cooling dq/dr must be of order q/H, and therefore in general large compared with q/r. If R be the radius of the earth, we see that the first term in this equation is of order $\lambda q/H^2$, the second $\lambda q/R^2$, and the fourth $g\rho_0 q/R$. Accordingly no term, with the exception of the third, can amount to more than a hundredth of the first. We can accordingly reduce the equation to

$$\frac{d}{dr}\{(\lambda + 2\mu)\,\delta\} - \frac{d\gamma}{dr} = 0 \qquad \text{.........(8)},$$

giving

$$\delta = \frac{\gamma}{\lambda + 2\mu} + \text{const.} \qquad \text{.........(9)}.$$

If we consider a point on the axis of x, the additional stresses are

$$p_{xx} = \lambda\delta + 2\mu \frac{dq}{dr} - \gamma \qquad \text{.........(10)},$$

$$p_{yy} = p_{zz} = \lambda\delta + 2\mu \frac{q}{r} - \gamma \qquad \text{.........(11)},$$

$$p_{yz} = p_{xz} = p_{xy} = 0 \qquad \text{.........(12)}.$$

Thus the principal stresses are radial and tangential, and it is at once seen from symmetry that this must be true at all points. Now the tendency of the matter to flow or fracture is determined by the stress-difference. As the initial stress-difference was zero, the actual stress-difference is the same as the difference between the radial and tangential additional stresses. If P, the radial stress, is the greater, horizontal compressional movement will tend to occur; if Q be the greater, there may be vertical fractures, or the matter may spread out horizontally. Now

$$P - Q = 2\mu\left(\frac{dq}{dr} - \frac{q}{r}\right) = 2\mu\left(\delta - \frac{3}{r^3}\int_0^r r^2\delta dr\right)$$

$$= \frac{2\mu}{r^3}\int_0^r r^3\frac{d\delta}{dr}dr = \frac{2\mu}{r^3}\int_0^r r^3 \frac{d}{dr}\frac{(3\lambda + 2\mu)\,nV}{\lambda + 2\mu}dr \qquad \text{.........(13)}.$$

Put
$$\frac{(3\lambda + 2\mu)\,nV}{\lambda + 2\mu} = \theta.$$

Then
$$P - Q = \frac{2\mu}{r^3}\int_0^r r^3\frac{d\theta}{dr}dr \qquad \text{.........(14)}.$$

Suppose first that the last adjustment to stress took place at solidification. Then the cooling is greatest at the surface, and steadily becomes less inwards. If λ/μ is nearly constant, as is usually true, $d\theta/dr$ is always negative, and therefore $P - Q$ is always negative. Thus the immediate effect of the cooling of the earth is to produce a strong tendency to vertical fracture or horizontal spread at all depths.

On the other hand, suppose that the crust has by flow or fracture adjusted itself since solidification until the horizontal tension first produced has been completely relieved, and consider the effect of further cooling. The fall of temperature at the surface is zero, on the supposition that the temperature there is maintained wholly by radiation from the sun, which is supposed constant. Hence θ is zero when $r = 0$, falls to a maximum negative value in the crust, and then increases again, reaching zero once more at the surface. Thus if it were not for the variation of r^3 within the region of integration, $P - Q$ would be zero at the surface. But above the layer of greatest cooling r is greater than below, and therefore in the integral $d\theta/dr$ is multiplied by a larger quantity when it is positive than when it is negative. Thus $P - Q$ is positive when $r = R$, and of order $R^2H\theta_1$, where θ_1 is the numerically greatest value of θ. At depths comparable with that of the level of greatest cooling, the integral is of course negative. Hence if there has been no variation in the surface temperature since hydrostatic

conditions were last attained, symmetrical cooling must necessarily lead to $P - Q$ being positive at the surface and negative below.

The radial stress P can be evaluated most easily by considering the equilibrium of a small piece within a cone of small angle 2α and between r and $r + dr$. The solid angle subtended at the centre is $\pi\alpha^2$, and the perimeter is $2\pi r\alpha$. The horizontal stress Q acts normally to the conical surface, and therefore produces a component force parallel to the axis of the cone. The equation of equilibrium is

$$\frac{d}{dr}(r^2P)\,\pi\alpha^2 dr - g\rho\pi r^2\alpha^2 dr - Q\,.\,2\pi r\alpha dr\,.\,\alpha = 0 \quad \text{.........(15)},$$

which gives easily

$$\frac{dP}{dr} = -\frac{2(P-Q)}{r} + g\rho \quad \text{.........(16)}.$$

Now $P - Q$ cannot exceed the breaking stress, say 10^9 c.g.s. units, and the first term on the right is at most 3 c.g.s. units. The second term is about 3000 in the same units. The deformation therefore makes no difference of importance to the vertical stress, which is practically the same as in hydrostatic equilibrium. It vanishes at the outer boundary, and with increasing depth (decreasing r) it becomes negative, corresponding to a pressure and not a tension. The type of yield involved in the contraction theory is in fact a crushing due to the horizontal pressure being greater than the vertical one; both are compressive stresses, and there is no question of the outer layers becoming separated from the inner ones. It is for this reason, incidentally, that the crushing strengths of materials are more interesting in problems of crustal mechanics than the tensile strengths.

15·35. With regard to the objections to the thermal contraction theory in particular, the best known is that of Osmond Fisher*, who claimed that the amount of contraction predicted was quantitatively insufficient to account for known mountain ranges. Osmond Fisher's discussion, however, rests on several incorrect hypotheses. In the first place, he uses Kelvin's theory of the cooling of the earth, which is now known to be in serious error, since it ignores radioactivity and takes the age of the earth to be only 10^8 years instead of over 10^9 years. This in itself shows that none of Fisher's quantitative estimates can be accepted at the present day. There is, however, a still more serious error. Fisher's criterion of quantitative accuracy is based, not on the reduction of area by crumpling, but on the volume of crumpled rock. This is estimated, correctly in principle, by finding the reduction of area in each layer down to the level of no strain, where crumpling ceases, and integrating with regard to the depth. Fisher then supposes the crumpled rock spread in a uniform layer over the surface of the earth, and finds that the thickness of this layer is only a few metres, which is small in comparison with the heights of existing

* *Physics of the Earth's Crust*, Macmillan, 1889.

mountains. With the data here employed, the depth found is greater than Fisher's estimate, being of the order of 200 metres, but is still less than the heights of known mountains. The comparison made, however, is quite illusory. It would be valid only if the physical process suggested in the calculation bore a close resemblance to that involved in the actual elevation of mountains, which is not the case. The rocks crumpled at great depths have not all been brought up to the surface in the process; a very small fraction of them have. Those crumpled at the surface have not been uniformly spread out, and if they had been, the surface would have remained perfectly level, and no mountains at all could have been formed. The only way of altering Fisher's comparison so as to make it serve as a trustworthy test of the thermal contraction theory of mountain building would be to find the depth of the layer that the existing mountains would form if they were all pulverized and spread uniformly over the earth; and it is certain that the depth of such a layer would not exceed a few tens of metres.

15·36. An objection of less fundamental character, but less easily answered, has recently been offered by Holmes*. The depth reached by cooling increases nearly in proportion to the square root of the time, and therefore the compression should do the same. The compression should therefore increase more and more slowly as time goes on, and the intervals between consecutive great upheavals should increase like the square root of the time. Holmes considers that the actual intervals are approximately equal, and the number in pre-Cambrian time does seem to be less than would be expected from the interval between the Hercynian (late Carboniferous) and Alpine folds, say 200 million years. There are two possible answers to the argument, though neither of them is so definite as one would like. The first is that the theory is being pushed into greater detail than a theory expressly constructed to deal with average conditions could be expected to stand when applied to the earth's surface, diversified as it actually is. The yielding in an epoch of mountain formation is not complete; it merely continues until the stresses are nowhere as great as the strength of the materials. Some stress-differences would therefore survive and be carried over till the next epoch. This effect would tend to distribute the mountain formation more uniformly in time.

But in one form of the thermal contraction theory approximate uniformity is to be expected. If the crust, when it became stiff, went into the liquevitreous state, further cooling from the top would produce convection currents. We should then have a durovitreous or crystalline layer on top, resting on a liquevitreous layer. The former thickens gradually, while the latter is still cooling by convection as fast as the slow conduction through the outer one will permit. If this is the actual state of affairs, the

* *Geol. Mag.* **64**, 1927, 274–275.

compression would vary at first as the square root of the time, but later the thickening of the crust, at an approximately uniform rate, would provide most of the compression*. This view of the process depends on several factors that have not yet been properly investigated, and is mentioned here merely to show that Holmes's objection is not final.

15·4. *Effects of Change in the Earth's Rotation.* We have seen that a simple calculation based on the present rate of secular acceleration of the moon's motion implies that 1600 million years ago the earth rotated in 0·84 of our present day, and that just after the moon was formed the period of rotation was probably about 5 hours. The ellipticity of figure is proportional to the square of the angular velocity of rotation, so that, on the above estimates, the ellipticity 1600 million years ago was about $\frac{1}{210}$, and the original ellipticity about $\frac{1}{13}$. Now the radius of the earth in colatitude θ is $R\{1 + \epsilon(\frac{1}{3} - \cos^2\theta)\}$. We are not at present considering variations of the mean radius, but it appears that the equatorial radius contains a variable part $\frac{1}{3}\epsilon R$. This would imply a contraction of the equatorial circumference by 18 km. in the last 1600 million years, and by 1000 km. since the moon was formed. The latter estimate is very large, and changes in the earth's ellipticity can hardly have failed to produce a marked effect on the primitive form of the earth's surface. On the other hand the compression in geological time is not very important. But if we consider that the moon was formed from the earth not more than 2500 million years ago we shall have to suppose the lags in the tides greater in the past, and the change of ellipticity in geological time greater than we have just found. A larger fraction of the available 1000 km. may be brought into geological time in this way. On the other hand the effect in polar regions is a stretching. The rate of tidal friction containing the inverse sixth power of the moon's distance as a factor, its effects must in any case have been concentrated in the early history of the earth, and mountains as we know them can probably not be attributed to it.

In addition to the change of ellipticity produced by the change of rotation, there is a general symmetrical contraction, noticed by J. W. Evans and evaluated by Stoneley†. This would have shortened every great circle by about 70 km. since the earliest times. But this effect is inseparably connected with the much larger effect of change of ellipticity, and the total effect, with its stretching of the crust over the great polar caps, is qualitatively different from the one that arises in our immediate problem. It appears that change of rotation, whatever its influence may have been, is unlikely to have determined directly the formation of existing mountains.

15·41. To sum up, the successes of the thermal contraction theory are that it predicts an amount of mountain formation of the right order of

* Cf. *Gerlands Beiträge*, **18**, 1927, 12–14. † *M.N.R.A.S. Geoph. Suppl.* **1**, 1924, 149–155.

magnitude, that it implies an alternation of quiescent periods with periods of mountain formation, with about the right number of each in the history of the earth, and that it accounts naturally for the existence of the Pacific type of mountain. In connexion with it, it must be recalled that it is not a theory invented *ad hoc* to explain mountain formation, but a direct consequence of a theory of the earth's thermal history designed primarily to account for the observed vertical variation of temperature in the earth's crust, and itself checked at several other points. None of the alleged objections to it appears to carry any weight except possibly that it does not fit immediately the observed lengths of the periods of quiescence, but even this does not appear insuperable.

15·5. *The History of Mountain Ranges.* We have seen that the level of no strain is at a depth of the order of 100 km. This is well down in the asthenosphere. Hence there is horizontal compression throughout the lithosphere and the upper part of the asthenosphere. On account of the weakness of the latter it must frequently give way, and in particular it must yield easily to the stresses developed before and during mountain formation. The outward displacement of the asthenosphere must go on while the mountains are being formed, and probably keeps pace with it, so that the great ranges are compensated from the start; it is not correct to speak of them as having been formed first and compensated afterwards.

15·51. But another type of outward flow is indicated by the later history of mountain ranges. We have already supposed that about 3 km. of sediments have been removed from Alpine valleys by denudation since the original compression, which took place about 30 million years ago. If denudation had proceeded at this rate since the Carboniferous period, say 250 million years ago, 25 km. of material would have been removed. This would be enough to remove the sedimentary and granitic layers, and to cut deeply into the intermediate layer. Even if we allow the upper layers to have been doubled in thickness by the compression, the depth of the bottom of the granitic layer below the outer surface would be only about 25 km. The actual conditions in very old mountain ranges are quite different. The mountains of Wales and Scotland, the Appalachians, and the Urals may be taken as typical. Old rocks, including gneisses that may have been derived from the granitic layer, are indeed usually exposed in the centres of such ranges; but the extraordinary fact is that old sedimentary rocks survive, even though they must have borne the full effect of the denudation. This seems to show that the average rate of denudation cannot have been more than a small fraction of that appropriate to Tertiary and Quaternary times, even when the regions compared are mountainous in both cases.

It seems unlikely that climatic conditions can have varied very greatly. The intensity of the atmospheric circulation is determined mainly by solar

radiation and dynamical conditions, which cannot have varied much, and if a mountain range is in the way of a current of moist air of given intensity the rainfall also will probably be about the same. We have no reason to suppose that the rainfall on a mountain range of given height has undergone great systematic variations, and if we do not make this hypothesis we must suppose that the real height of the old mountain ranges, during most of the time since their formation, has been much less than that of the Alps, and that denudation has been small because the relief has been low. This view is supported by the fact that the old ranges in question are now of moderate height, though the amount of compression indicated by the degree of folding is comparable with that in the greater ranges.

A further argument, not depending on assumptions about the rate of denudation, is provided by the fact that some old mountain ranges showing great folding have been denuded so far as to be almost indistinguishable from plains. How can a mountain range be so much lowered? We have seen that the removal of 3 km. of sediments lowers the outer surface by only about 0·8 km. Thus if a mountain range was originally 3 km. in height, and the sediments composing it were 3 km. thick, it could not be lowered to the level of the surrounding country until first the 3 km. of sediments, and then 11 km. of granite, had been removed. Here again we have a result inconsistent with the presence of old sediments in the ranges in question.

The only way of reconciling these facts seems to be to admit that denudation, accompanied by isostatic compensation, is not the only cause of the lowering of mountains. If the outer layers can move outwards under the excess of pressure due to the weight of the mountains, the outer surface can be lowered without heavy denudation. If such outflow reduced the thickness of the granitic layer by 10 km., the outer surface would be lowered by 2 km. If the intermediate layer is tachylyte and became 20 km. thinner, a similar result would follow. Outflow in the lower layer of course would not account for anything, since nearly complete adjustment in this layer is already implied by the isostatic compensation of the recent ranges; it can do nothing more for the old ones than it has already done for the new ones. The mechanism is to be sought in the granitic and intermediate layers.

The normal temperature being nearer to the melting point in the intermediate layer than in any part of the granitic layer, we naturally suppose that most of the outflow takes place in the intermediate layer. In the history of a mountain range we therefore contemplate the following stages.

1. Compression of the crust, with thickening of the sedimentary, granitic, and intermediate layers, elevation of the surface and outflow in the lower layer.

2. Sculpturing of the outer surface by denudation, with deepening of the valleys and slight rise of the highest ground. This process leads to the maximum of relief.

3. Outflow of matter from the intermediate layer, with compensating inflow in the lower layer. The region sinks, and denudation becomes less active. The valleys are by this time so deepened as to undergo no further denudation, and further erosion is confined to the mountain tops.

4. Final stage, when the whole is worn down. Old sedimentary rocks and possibly the granitic layer are exposed. The granitic layer is probably thicker, and the intermediate one a great deal thinner, than at the commencement of the process.

The time taken in the first stage is of the order of a million years, the second 30 million, and the third 300 million.

15·52. These processes are phenomena of plasticity, the quantitative theory of which has not yet been far developed. If viscosity is measured as the ratio of shearing stress to rate of shear, it is always infinite when the stress-difference is below the strength. When distortion is actually taking place, however, the viscosity is finite. The best we can do at present is to use the elasticoviscous law of 14·422. Under forces of long period an elasticoviscous solid with rigidity μ behaves practically as a liquid with viscosity μt_1. For a mountain range 100 km. wide to be flattened out in an interval between 10 and 300 million years, it can be shown that a kinematic viscosity between 10^{22} and 3×10^{23} cm.^{2}/sec. in the intermediate layer will suffice. Such a viscosity in the lower layer, where flow is not constricted in depth, would imply adjustment in a time not exceeding 3×10^4 years*. This can be regarded as the time needed for ordinary isostatic compensation. It is of the same order as the time since the last glacial period, and indicates that most, but perhaps not all, of the inequalities left by it have had time to become compensated†; and it confirms our preliminary conjecture that mountains are compensated nearly as fast as they are formed.

The viscosity in the lower layer needed to prevent convection currents in the uppermost 300 km. in spite of the gradient of temperature is of order 5×10^{22} cm.^{2}/sec. That indicated by the variation of latitude for the rocky shell in general was at least 5×10^{20} cm.^{2}/sec. But it should be noticed that damping of the variation of latitude, and also convection currents, would be absolutely prevented if the material has a finite strength, however small, and it probably has one. The estimates from the history of mountains refer to the behaviour under stresses great enough to cause permanent deformation, and give positive evidence about the imperfection of elasticity of the crust.

15·6. *Vulcanism and Intrusion.* The formation of volcanoes, dikes,

* *Gerlands Beiträge*, **18**, 1927, 22–25.

† Nansen gives evidence for the view that the Scandinavian ice sheet had time to become compensated (*loc. cit.*). Cf. also Daly, *Our Mobile Earth*, 190–202; Lake, *Geog. Journ.* **72**, 1928, 575–576.

and sills, and other elevations of rock magma, in a liquid or partly liquid state, to the surface or near it require explanation. On the theory of Chapter VIII the present temperature at the base of the intermediate layer is about 650° C., and that at the base of the granitic one about 300°. Dry basalt and granite do not soften at temperatures below 1000° or more, and this value may have to be increased at the depths in question to allow for the effect of pressure on the melting point or the viscosity. The generation of liquid granite and basalt is therefore a definite problem; so definite, indeed, that it formed Holmes's principal reason for abandoning the theory of the earth's thermal history that he had played a leading part in creating. I do not, however, think that the difficulty is so serious. In the first place, we cannot admit that fusion in the upper layers is a normal condition. Vulcanism is occasional and local, not perpetual and worldwide, and the behaviour of the earth in supporting mountains, transmitting distortional waves, and permitting oceanic tides shows quite definitely that the upper layers are solid at present. Secondly, Jaggar's observations on the crater of Kilauea* showed that the basaltic magma there was truly fluid at temperatures of 750°–850°, the reduced temperature being due to volatile constituents. Elevation of magma to the surface being the abnormal condition, we must suppose the normal temperatures in the basaltic layer to be below that of newly risen magma. From this point of view the temperatures inferred for the intermediate layer seem highly satisfactory. Indeed it seems to me that the real difficulty is in the opposite direction to that felt by Holmes. Direct measures of the viscosity of rocks at high temperatures have not been carried out, but Trouton and Andrews† found that soda glass at 710° had a viscosity of 4×10^{10}, and at 575° one of $1{\cdot}1 \times 10^{13}$. Extrapolation to 500° gives a viscosity of 2×10^{14}. We have seen that the kinematic viscosity of the intermediate layer is over 10^{22}, and the ordinary viscosity over 3×10^{22}. It is rather surprising, in consideration of the temperature, that the viscosity should be so high, but the legitimacy of the comparison needs to be checked experimentally.

15·61. Our problem is to explain how the basaltic layer (we are assuming for the purpose of discussion that the intermediate layer is tachylyte) comes to be heated up by the requisite 100° at least to make it capable of rising to the surface through the granitic and sedimentary layers, and how, when it has done so, it manages to solidify again. The great basaltic outpours, especially the plateau basalts, are mainly in districts where active volcanoes no longer exist. These facts seem to point to local and temporary sources of heat. So does the difference in level of the lava in two neighbouring craters in Hawaii ‡. Chemical action is capable of producing elevations of temperature of the correct order of magnitude,

* *Amer. J. Sci.* (4) 44, 1917, 208–220.

† *Phil. Mag.* 7, 1904, 347–355.

‡ Cf. Daly, *loc. cit.* 164–165.

for Jaggar shows in the paper just quoted that the lava in the fountains of the Kilauea crater is at about 1100°, though variable; the normal temperature some distance down was 750°–850°, and the difference was attributed to chemical action with the atmosphere. Day has gone a long way towards a constructive theory of vulcanism*, in which chemical action between the gases expelled from the crust plays a leading part. The gases actually emitted at the surface in volcanoes are such as would react mutually if they were mixed, and it is certain that they must meet within the crust. Granting the facts, the explanation of local fusion when active gases from different sources in the lower layer react in the intermediate layer seems adequate. It has the further recommendation that when a given source of gas is exhausted or blocked up the fusion will cease, and the igneous activity come to an end.

15·62. Mountainous regions would evidently be expected to be regions of igneous activity. If the formation of a mountain chain thickens the upper layers locally by one-half, the ultimate effect when conditions have again become steady is to increase the local output of radioactive heat in the same ratio, while the basal temperature S_0 of Chapter VIII is multiplied by 2·25. The former change is capable of observational test, but the existing data suffer from ambiguities of interpretation. The latter implies that fusion in the intermediate layer, and possibly even the granitic layer, is to be expected below a mountain chain. The squeezing out of the intermediate layer already inferred, and the occurrence of large intrusions of granite and andesite in mountainous regions, are therefore not surprising. The granite in these cases usually forms a bathylith or central core, and andesite is the rock usually associated with volcanoes among folded mountains. There is, however, a difficulty in seeing how igneous activity produced by radioactive heating can stop, because such heating is steady, and steady heating in a solid or liquid normally leads to a distribution of temperature tending asymptotically to a steady state such that conduction and convection remove the new heat as fast as it is generated. Fusion once produced would therefore be permanent. The tendency of radioactive constituents to concentrate upwards seems to save the situation here once more, though the thinning of the sedimentary and granitic layers by denudation, and of the intermediate one by horizontal outflow, will help.

15·63. Excess of radioactivity to such an extent that S_0 exceeds the fusion point in general, and not merely in the special conditions of mountain ranges, has been appealed to as an explanation of the igneous activity in non-mountainous regions. In this case the usual difficulty of seeing how resolidification can ever occur is aggravated. In such regions as the Deccan, for instance, where basaltic outpours have covered the outer surface over

* *J. Frank. Inst.* 200, 1925, 161–182.

wide areas, the total radioactivity in a vertical column is unaffected by the redistribution, but the highly radioactive granite layer is buried to a greater depth than usual, with corresponding further increase of S_0. In these circumstances it looks as if the outbreak must continue until the whole intermediate layer has poured out and submerged the outer surface to a depth of 20 km., followed by fusion of the granitic layer and upward concentration of radioactive matter. (We may recall at this stage that plateau basalts are *less* radioactive than ordinary basalts.) It does not appear that any of the great outbreaks of igneous activity have been on anything like the scale to be expected had they been due to the excess of radioactive heating. On the other hand Day's chemical hypothesis appears to fit the facts quite well; it is merely necessary to assume enough temporary heating to melt the basalt locally and produce ejection of part of the magma; the resulting submergence of the outer crust would increase S_0, but there is no reason to suppose it increased enough to produce permanent fusion, and the outbreak would cease when the supply of chemically active material ceased.

15·64. *Vulcanism on the Moon.* The mean density of the moon is 3·4. We have seen that the effect of compression in the earth's rocky shell is enough to increase the mean density from 3·3 to 4·3. Gravity on the moon is only a fifth of what it is on the earth, and the radius of the moon is about half the depth of the rocky shell. Thus the effect of compression on the mean density in the moon should be of the order of a tenth of what it is in the earth's rocky shell. The moon is therefore probably mainly composed of material with a density of about 3·3 at atmospheric pressure. There seems to be no room for a metallic core in the moon. On the other hand the opinion that the earth's rocky shell is mainly dunite and that the moon has been in some way formed from the outside of the earth is in good agreement with the moon's density. In cooling the moon would be expected to have gone through a series of stages analogous to those inferred for the earth. The separation of granitic and intermediate layers, possibly rather thicker than on the earth, since the moon came from the outside of the earth, and the slow upward concentration of radioactive matter, assisted by the upward movement of gases, especially steam, would proceed as for the earth. On the earth the visible traces of this process have long been obliterated by denudation, but on the moon the indications of a former spell of volcanic activity incomparably more violent than anything known on the earth are clear. The great smooth plains (maria) cover a large fraction of the surface, and the lighter-coloured remainder is largely composed of craters of widely different sizes. In places, however, there are stretches of elevated ground of very irregular topography, but without craters*. It looks as if the latter represent the

* A remarkably fine collection of illustrations of lunar formations is given by W. Goodacre in Hutchinson's *Splendour of the Heavens.*

primitive solid surface, through which the eruptions took place during the primitive expulsion of the water from the interior, forming the majority of the craters*. The maria are a further set of outpourings on a much wider scale, and appear to have submerged much of the original surface. Minor activity has taken place more recently, as is shown by the few craters existing within the maria. The surfaces of the maria are not perfectly flat, but the departures are not greater than can reasonably be attributed to the changes of volume of a viscous liquid during cooling. The moon's apparent complete quiescence at present, in spite of its riotous early history, is explicable on the supposition that in it, as in the earth, gradual upward concentration of the radioactive constituents once played a determining part in its development. There is, however, a difference. Whereas in the earth this concentration overshot the mark and reached a stage that permitted steady cooling afterwards, in the moon it must have stopped earlier: there are no folded mountain ranges on the moon†.

15·7. *The Origin of Continents and Ocean Basins.* Two things are clear concerning the initiation of the distinction between continents and oceans. It did not take place while the earth was fluid, for in the fluid state surfaces of equal density would remain equipotentials, and the free surface would be perfectly smooth. It cannot have occurred since the earth solidified, for the process of collecting nearly all the granite into one-third of the earth's surface seems inconceivable when it is resisted by the rigidity of the materials. We seem to be restricted to processes occurring during solidification.

The best known of these is based on the resonance theory of the origin of the moon. If the moon was formed by this means while the earth had a thin solid crust, mainly granitic, it would have carried off a large portion of the crust with it. The hole that remained would quickly become partly filled by inflow of heavy magma, leaving large granitic slabs floating on a heavy fluid. These slabs, according to this theory, became the continents, and the exposed denser material the ocean floors. The place where the moon emerged gave the largest ocean floor; the Pacific Ocean is thus regarded as the scar left when the moon was formed.

This theory, due to Osmond Fisher, has had many adventures since it was formulated, partly through the parent resonance theory and partly on its own account. The difficulty about the resonance theory has always been to keep enough of the earth with a low viscosity for a long enough time for a displacement with the requisite amplitude to be worked up. But we have now seen that only a thin shell of the earth could have

* Explosions of magnesium powder and potassium chlorate under a plastic surface have been found by S. Mohorovičić to give formations remarkably like lunar craters: *Arhiv z. Hemiju i Farmaciju, Zagreb*, **2**, 1928, 66–76.

† Unless indeed the elevated irregular regions are what our mountains would have been like if they had not been denuded.

solidified before the radioactive materials became concentrated to the top; there is no clue as yet to the time taken in this stage, and so long as it lasted the earth would behave practically as a perfect fluid in tidal disturbances. There is therefore no serious objection to the resonance theory now. An objection expressed to Fisher's theory in the first edition of this work has proved to depend on a misunderstanding of a result proved by Jeans. Light solid fragments floating on a liquid interior, and largely confined to one side, would correspond to a first harmonic displacement. Now Jeans showed* that the earth is stable as regards displacements of this type. This appears to show that the fragments would spread out so as to distribute themselves as symmetrically as possible, so that the scar would have lasted a few days or months instead of to the present time. But Jeans's argument starts with a symmetrical earth, and is no longer applicable when the cohesion of each floating fragment is resisting the restoration of the spherical form instead of the disturbance from one. It actually appears that the force between two floating blocks is an attraction, so that the primitive continents would tend to collect together; but it is small and probably insignificant in comparison with another force making for horizontal displacement, the Polflucht of Eötvös†.

The theories of the two forces are very similar. Imagine a block of depth h and density ρ floating in a liquid of density ρ_0. The base of the block is at a depth $\rho h/\rho_0$ below the surface of the fluid; the height of its centre of mass above that of the displaced fluid is $\frac{1}{2}(1 - \rho/\rho_0)\, h$. If the block is removed to a great distance and replaced by an equal mass of the fluid brought from a great distance, the new centre of mass is that of the displaced fluid, and the work done by gravity is $\frac{1}{2}(1 - \rho/\rho_0)\, mgh$, where m is the mass. This is the potential energy of the system with the floating block, in comparison with a fluid mass in equilibrium. If g is not the same all over, we see that the potential energy is least if the block is where gravity is least; there is therefore a force tending to displace a floating body towards the regions where gravity is least. Thus a floating body tends to move towards the equator; this is Eötvös's phenomenon. With the formula for gravity

$$g = G\,(1 + 0{\cdot}00529 \cos^2\theta),$$

the equatorward force per unit area is

$$\left(1 - \frac{\rho}{\rho_0}\right)\rho\,\frac{h^2}{R}\,G \times {\cdot}00529 \cos\theta \sin\theta.$$

With

$\rho = 2{\cdot}7$, $\rho_0 = 3{\cdot}3$, $h = 15$ km., $R = 6{\cdot}4 \times 10^8$ cm., $G = 980$ cm./sec.^{2},

the maximum value of this is about 4400 dynes/cm.^{2}

* *Proc. Roy. Soc.* A, **93**, 1917, 413–417. The argument of 12·33 contains an alternative proof.

† *Verh. d. 17. Konf. d. Int. Erdmessung*, 1913; W. D. Lambert, *Am. J. Sci.* **2**, 1921, 129–158.

The attraction between continents arises from the disturbance of gravity at the centre of one produced by the other. It is much less than the equatorial force. The thickness of a continental block is comparable with the height of the equatorial protuberance, but its disturbing effect is much less on account of the fact that it is compensated. The force per unit surface between average continents 90° apart is of the order of 10^{-2} dynes/cm.^{2}

If we attribute to the earth a viscosity of 10^6, such as we have seen to be permissible, and allow for the greater value of the equatorial force owing to the greater speed of rotation in early times, it is found that a floating continent would reach the equator in a time of the order of a day*. It is therefore difficult to see how continents formed in this manner could have failed to become nearly symmetrically distributed across the equator and stay there. Nevertheless the difficulty is less serious than the previous one, and it does not seem that Fisher's theory of the origin of the Pacific is impossible.

15·71. An alternative theory, not depending on the birth of the moon, is based on the result of 15·34 that the first effect of cooling from the surface is to produce a horizontal tension at all depths. This would tend to produce vertical cracks, of small horizontal scale to begin with, which would become the localities at first of formations analogous to lunar craters. As the process went on, these cracks would tend to spread, on the principle that it is easier to extend an old crack than to start a new one. Also the deeper cracks would tend to grow downwards faster than the shallower ones, and the cracks would tend to run together at their ends, on account of the region between them having to support more than its share of the tension. The honeycomb of small cracks would therefore tend to be dominated by a few large polygonal systems. Such a process is indicated by the formation of the lunar maria after the majority of the craters. But a tension in a crust too small to fit the interior would tend to expel through the cracks enough fluid to relieve the tension. On reaching the outside this would spread out horizontally, but it would be delayed by viscosity and might well have failed to overflow the whole surface before it solidified. If this happened when the cracks had not penetrated through the granitic layer the ejected magma would be mainly granitic, but if it happened later it might be basic or ultrabasic. The apparent total absence of granite from the Pacific is against this theory, because the theory requires at least a thin solid layer all over the outside before the formation of the fissures; but Holmes's suggestion of a syenite layer over the ocean floor is at least helpful. It would apparently make the continents on the earth correspond to the maria on the moon, which is curious. F. E. Wright has noticed

* *M.N.R.A.S. Geoph. Suppl.* 1, 1926, 418–423.

that the light reflected by the moon shows remarkably little polarization*, indicating the absence of both glasses and basalts from the surface, but does not distinguish between the maria and the elevated regions. Further work on these lines is desirable.

Mars shows a diversity of surface markings, and as it does not possess a large satellite Fisher's theory is not available as an explanation. That just given may be; but the origin of continents cannot at present be considered a solved problem.

15·8. *The Permanence of the Continents.* Three questions are really involved under this head, and require separate treatment. Has the inclination of the earth's axis to the plane of its orbit varied during its history? Have the poles always been the same points of the outer surface? Has the outer surface itself undergone considerable distortion, other than the recognized displacements involved in mountain formation and other geological processes?

15·81. The answer to the first question is definitely 'Yes.' The theory of tidal friction expounded in Chapter XIV is a simplified one, and assumes the equator and the planes of the earth's and moon's orbits to coincide. The fact that they do not introduces several additional tides depending on the inclinations and eccentricities of the orbits, and all these are affected by friction. The investigation of their effects formed the subject matter of most of the second volume of Darwin's *Scientific Papers*, and he was able to show how they could have brought the system to its present state from an initial one with the earth and moon close together, the moon's orbit nearly circular, and in the plane of the equator, and the equator at any rate much less inclined to that of the ecliptic than it is now. But he was using the hypothesis that the tidal friction is bodily, and it is not yet known how far his results would be altered if dissipation in strong tidal currents in shallow seas is the dominating type. Even the direction of the changes is somewhat in doubt, though the problem is probably capable of a definite solution.

In any case the effects of tidal friction on the inclinations must be slow and have acted in the same direction through most of the time concerned. In addition there are various periodic effects arising from astronomical causes, but all of these are small and have short periods, judged by geological standards. A further influence arises from internal changes in the earth, but is not important. If we consider the axis of the earth's angular momentum, this can change in direction only through couples acting on the earth from outside. Apart from the tidal effect, the only couples concerned are due to the attractions of the sun and moon on the equatorial protuberance. Their main effect is the precession of the equi-

* *Proc. Nat. Acad. Sci.* **13, 1927, 535–540.**

noxes, a motion of the axis of spin in a circular cone about the pole of the ecliptic; the associated changes in the inclination are small and have short periods. The inclination of the axis of the earth's angular momentum to the plane of the orbit can have changed appreciably only through tidal friction. Further, the difference between the axis of angular momentum and the axis of rotation is very small, depending only on the actual angular momentum of the horizontal movements in progress at any time, and on the variation of latitude*.

15·82. The other two questions do not concern the movement of the earth as a whole, but only the displacements of parts of it relative to one another. They depend on a discussion of what forces exist and what displacements they are capable of producing. The chief of these deforming forces are the equatorial attraction, tidal friction, the attraction between continents, and perhaps an internal stress incidental to the precession of the equinoxes. The magnitude of the first is, we have seen, of order 4000 dynes/cm.^{2} Tidal currents at their strongest give a bottom drag of the order of 40 dynes/cm.^{2}; but this is abnormal and is reversed in direction in every tide where it does occur. The mean secular tidal friction producing the slowing down of the earth's rotation corresponds to a westward stress of the order of only 10^{-4} dynes/cm.^{2} over the earth's surface. The attraction between continents may, we have seen, give stresses of the order of 10^{-2} dynes/cm.^{2} The stress in precession that makes the whole of the earth precess at the same rate, instead of different shells precessing at different rates, is at most of order 60 dynes/cm.^{2}, and this must be mainly alternating in direction. The chief of these factors is definitely the equatorial attraction. Now a horizontal stress of 4000 dynes/cm.^{2} over a circular continent of radius 2000 km. would give a total force of $4000\pi\,(2 \times 10^8)^2$ dynes. This is resisted by the strength of the lithosphere, say 40 km. thick. An average stress of 10^5 dynes/cm.^{2} around a cylinder of radius 2000 km. through the lithosphere would suffice to prevent all movement. The strength at the surface is of order 10^9 dynes/cm.^{2}, and the support of the great mountains seems to point to greater strengths lower down. There is therefore not the slightest reason to believe that bodily displacements of continents through the lithosphere are possible.

Even the strength of the order of 10^7 dynes/cm.^{2} inferred for the asthenosphere, on several grounds, would be enough to prevent permanent distortion by forces of the magnitude suggested; and it may be well to recall the evidence of the moon's ellipticity that this strength is applicable to stresses lasting for intervals of the order of the whole age of the earth.

To adopt the view that the earth's crust has no strength under stresses of long duration, in spite of the absence of theoretical justification for such a view and the definite evidence against it, makes it possible to account

* A full discussion is given in Darwin's *Papers*, **3**, 1–46.

for some continental displacement, but not much. In this case the deforming stresses produce motion, resisted by the viscosity of the material undergoing deformation. We have seen that an ordinary viscosity (μ) of 3×10^{22} c.g.s. is a lower limit to the viscosity of the intermediate layer under the stresses due to the weight of mountains. Under smaller stresses the viscosity may be higher. With this viscosity and a shearing stress of 10^5 dynes/cm.^{2} the rate of shear would be 3×10^{-18}/sec., and notable distortion would be produced in 3×10^{17} sec. or 10^{10} years. But in the upper layer the viscosity is certainly much higher. The estimate just used is such as to permit a mountain chain to be flattened in 10^7 years. The actual time is perhaps 30 times as long. Further, the estimate refers to the intermediate layer; if the granitic layer had a similar viscosity the old mountain ranges would not be still standing. Thus the actual time needed to distort the outer crust is much more than 10^{10} years, and a large multiple of the time available. The question concerning the possibility of considerable distortion of the outer crust during geological time must therefore be answered in the negative.

15·83. Displacements of the outer crust bodily over the interior would involve no distortion of the outer crust except the small up and down ones rendered necessary by the ellipticity of figure. The elasticity and viscosity of the lithosphere would not resist this process. Now the variation of latitude shows that the viscosity of the rocky shell in general exceeds 5×10^{20} c.g.s. A shearing stress of 4000 dynes/cm.^{2} over the surface would produce shear at a rate $0{\cdot}8 \times 10^{-17}$/sec., and the crust would be displaced through a radian in a time of the order of 10^{17} sec. or 3000 million years. But this rate must be reduced to allow for the fact that this is the equatorial drift; the continents cannot move independently, but can only turn the crust as a whole so as to bring the greatest projections over the equator, so that the stress is resisted by the viscosity of the whole interior and not merely by the parts immediately under the main continents. It seems unlikely that this cause can have turned the crust by more than 5° during geological time.

15·84. The answers to our questions seem to be, therefore, as follows. A slow secular change of the inclination of the earth's axis to the ecliptic during geological time is to be expected; the change has probably been on the whole an increase, but this is not quite certain. Secular drift of continents relative to the rest of the crust, such as have been maintained by A. Wegener and others, are out of the question. A small drift of the crust as a whole over the interior is not impossible; it would be shown as a displacement of the poles relative to the earth's surface, but the maximum amount seems too small to be of much interest.

15·85. Vertical movements of the land surface, apart from mountain formation, are to be expected on several grounds. In a perfectly sym-

metrical earth the stresses developed in the intervals between epochs of mountain formation would not disturb symmetry, but in the actual earth the differences of structure from place to place, and especially the difference between continents and oceans, must make the purely elastic yielding asymmetrical, with the development of vertical movements of large horizontal extent. The same applies to the intermittent adjustment of the ellipticity to the value appropriate to the varying speed of rotation.

Land bridges connecting different continents have been proposed at intervals by geologists, largely to provide routes of migration for animals and plants. Their popularity appears to have diminished in recent years, in view of the alleged explanations of the same phenomena by the theory of continental drift. If ever there was a migration from the frying-pan into the fire it is this. The main objection to the theory of former land bridges, which have sunk below the sea, is an apparent conflict with isostasy, which would be serious if we were restricted to two materials each capable of only one physical state, for then the quantity of the lighter material per unit area would definitely determine the elevation of the land surface, and serious change in the height of the land on a continental scale would be very difficult to explain. But when we have three materials, probably each capable of a vitreous state and at least one crystalline one, the question is on a very different footing*. If for instance 20 km. of tachylyte crystallized to eclogite a depression of level of 3·6 km. without departure from isostasy would be the result; this is enough to account for the foundering of any land bridge that has been suggested. (It may be recalled that there is no seismological evidence of the existence of dolerite or gabbro as such within the crust.) In addition change of state in the lower layer seems to admit a great variety of hypotheses. But I do not actively advocate any of them, because I am not yet convinced of the cogency of the palaeontological evidence. All that is necessary to establish flora and fauna of similar character in two regions of similar climate and soil is a single migration. The speed of an average ocean current is about 1 knot, which would take a given floating object across the South Atlantic in a few months. The species concerned are plants and low animals, and it seems far from impossible that spores, seeds, or eggs could have drifted across on floating refuse. Such an occurrence would doubtless be rare, but as it would explain the phenomena without needing to happen more than a few times its rarity is no objection.

The depression of the crust by the weight of sediments is of some interest. If the sediments are deposited on land and have density 2·3, the effect of outflow of matter of density 3·3 below is to lower the surface by 0·70 km. for each kilometre of material added, leaving the actual surface

* It is for this reason that I think Suess's names 'Sal' (modified to 'Sial' by later writers) and 'Sima' have now outlived their usefulness.

raised by 0·3 km. Thus 3·3 km. of sediments could be deposited in a valley a kilometre deep before it was filled up.

If sediments of density ρ_1 are deposited from water of density ρ_2, and k is the depth of the sediment, x the depression of the original surface, and ρ_0 the density in the lower layer, the mass per unit area is increased by

$$\rho_1 k - (k - x)\rho_2 - \rho_0 x,$$

and the condition for compensation is that this shall vanish. Hence

$$k = \frac{\rho_0 - \rho_2}{\rho_0 - \rho_1}(k - x).$$

But $k - x$ is the reduction in the depth of the water, and therefore this equation gives the depth of sediments capable of being deposited in water of given depth. With our previous values, and $\rho_2 = 1$, the coefficient is 2·3; thus 2·3 km. of sediments can be deposited in water originally of depth 1 km.* This estimate may be too high if the deposition is over a small area, on account of the regional character of the compensation.

15·9. *Summary*. The effect of denudation on a newly formed mountain range, combined with compensation, is to raise the mountain tops at the expense of the valleys. Allowing for this, we can make an estimate of the crustal thickening, and hence of the amount of compression involved in the formation of the range. The amount found is comparable with that measured in the field, but in some cases, notably the Alps, is appreciably less, and suggests strongly that part of the compression indicated by the observed folding is only apparent and not due to crustal shortening. The compression predicted by the thermal contraction theory is of the right order of magnitude, so that thermal contraction is in any case a major cause of mountain formation, and is presumably the chief cause until some other is produced. The formation of the Pacific type of mountains, and the long intervals of quiescence, are natural consequences of the theory. Some alleged objections to the theory are discussed.

The lowering of old mountain ranges to moderate height or even to base level by denudation is shown to lead to other consequences disagreeing with observation unless we assume that squeezing out of the intermediate layer under the weight of the mountains plays an important part in lowering them. Estimates of the viscosity within the crust, based on the time scale in the evolution of a mountain range, are derived.

Igneous activity is discussed with reference to the influence of excess radioactivity and of chemically active volatile constituents. The origin of the moon's surface features and of the continents are considered; also the question of the permanence of the continents and of the inclination of the earth's axis to the ecliptic.

* A. Morley Davies, *Geol. Mag.* 1918, 125 and 233; E. M. Anderson, *ibid.* 192.

APPENDIX A

The Planetesimal Hypothesis

"The man who makes no mistakes does not usually make anything."
EDWARD J. PHELPS.

A·1. The Planetesimal Hypothesis was historically the parent of the Tidal Theory of the Origin of the Solar System, elaborated in Chapter II. It was invented by T. C. Chamberlin and F. R. Moulton in the early years of the present century, and detailed accounts of it may be found in Chamberlin and Salisbury's *Textbook of Geology*, Moulton's *Introduction to Astronomy*, and Chamberlin's *The Origin of the Earth*. Like the theory here adopted, it supposes the sun to have been broken up by a passing star, and there is a general resemblance between the modes of formation and rupture of the filament on the two theories. The authors of the Planetesimal Theory suppose that two filaments were formed, projecting from the sun at diametrically opposite points. Jeans has shown that it is possible that only one filament was formed; the theory here developed works just as well with only one, and it has been suggested that it is possible that the shorter was wholly reabsorbed into the sun, even if it was ever formed. At this stage, however, the differences between the theories begin to become serious. The authors of the Planetesimal Theory believe that the planets would cool principally by adiabatic expansion, whereas it is shown here that at any rate the larger ones would cool principally by radiation from the surface. They assert further that all the planets, large and small, would form liquid drops at once, that these would quickly solidify, and that the planets formed by the aggregation of the solid particles would be themselves solid from the start. It has been shown in Chapter II that, in whatever way the planets cooled, they would always pass through a liquid stage.

A·11. *Impossibility of Great Accretion.* Perhaps the most serious divergence between the two theories is, however, in the nature of the postulated resisting medium. In the present work it is supposed to be a gas, probably consisting chiefly of hydrogen; in the Planetesimal Hypothesis it is supposed to be composed of particles that solidified during the condensation of the planets, but acquired velocities so great that gravity could not retain them. These would then revolve around the sun as independent bodies; they are the 'planetesimals' that give the theory its name. It is supposed that they were afterwards largely swept up by the planets, and that their effect was to reduce the eccentricities of the planetary orbits to their present values. Now it is possible, and indeed almost certain, that many such small solid particles were actually set in motion during the cooling of the primitive planets, but there is a grave objection to supposing that they can have had any important effect on the orbits of the planets. Their orbits, like those of the planets, must have been initially highly eccentric. The gravitation of the planets would make their apsidal lines rotate in at most some thousands of years, and thus, even though they might be moving without collisions initially, they would, in a short time cosmo-

gonically, reach a state such that any region large enough to contain a moderate number of them would contain nearly equal numbers moving inwards and outwards with velocities comparable with the velocity of a planet moving in a circular orbit in the same neighbourhood. Now if two solid bodies moving with such velocities collided, they would certainly be volatilized. Meteors, for instance, are volatilized when they enter the earth's atmosphere with such velocities, even in spite of the opportunity for loss of heat by convection and radiation during the several seconds the flight lasts. In the absence of air, the impacts of the bare surfaces would ensure that the whole of the energy loss on account of the imperfection of restitution would be liberated instantly, and thus, still more than in the case of meteors in the atmosphere, volatilization would ensue.

If now we suppose the planetesimals to be spheres, of radius c, the probability of any particular planetesimal striking another body, of radius a, is, for the same relative velocity, proportional to the square of the sum of the radii. Thus the probability of a planetesimal hitting any other particular planetesimal is to the probability of its hitting a planet of radius a in the ratio $4c^2/(a + c)^2$ or practically $4c^2/a^2$; and if we add up the probabilities for all planetesimals, the probability of a particular planetesimal hitting any other planetesimal in general is to that of its hitting a planet in the ratio of four times the surface of all the planetesimals to the surface of the planet. It can easily be seen that this result is still correct as regards order of magnitude when the planetesimals are no longer supposed spherical. The condition that more planetesimals were swept up by the planets than were volatilized is therefore that the total surface of the planetesimals was much less than that of the planets.

Now let us consider the effect of accretion on a planet moving in an eccentric orbit. When such a planet is near aphelion its velocity is less than that corresponding to a circular orbit at the same distance, while when it is near perihelion its velocity is greater than that for a circular orbit. If we suppose for the moment that the planetesimals were moving in circular orbits, the planet would therefore overtake them when near perihelion, and be overtaken by them when near aphelion. Now we know from the theory of impact that whenever two bodies unite into one the velocity of the combined body is between those of the original ones. Thus the difference between the velocities of the body and of the planetesimals in its neighbourhood would be diminished by every impact, and therefore its orbit would gradually become more nearly circular, just as in the case of a gaseous resisting medium. If we allow for the fact that the planetesimals were moving in highly eccentric orbits like those of the planets, the effect would be more marked, for a planet would pick up more particles it met than it overtook. We can therefore agree that the eccentricity would decrease. The authors of the Planetesimal Hypothesis say that the eccentricities would be reduced in this way to a fraction of their original values, and that the nearly circular orbits of the planets are explained. The explanation is, however, insufficient. If a planet picked up a planetesimal of its own mass, moving in a circular orbit, the difference between its velocity and that of a body describing a circular orbit would only be halved, and it is not difficult to show that the finely divided condition of the matter picked up does not alter this conclusion. Hence to produce a considerable reduction in the eccentricities of the planetary orbits the total mass of the planetesimals picked up must have been comparable with

the masses of the planets themselves*. But the more finely divided the matter the greater would its surface be, and therefore the total surface of the planetesimals must have exceeded many times that of the planets. This result, by the last paragraph, shows that collisions between planetesimals must have been enormously more frequent than those between planets, and therefore the planetesimals must have been volatilized by collisions among themselves before they had time to affect the eccentricities of the planetary orbits appreciably. Thus the reduction of the eccentricities of the orbits of the planets by accretion is impossible.

A·12. This discussion has constructive value, in that it shows that the masses of the planets cannot have increased by any considerable fractions of themselves since they were formed. Thus all the meteors picked up cannot have had much effect on the earth's size or on its orbit. The solid particles that left the planets in their early history may have collided and volatilized one another; in that case they would have been added to the gaseous resisting medium, and may in that form have had some cosmogonical importance.

A·2. *Possible Compression.* Though the impact of solid planetesimals on the surfaces of the planets cannot have had much effect on the masses and motions of the planets, it is as well to examine other arguments that have been advanced against the theory of the former fluidity of the earth and in favour of the view of Chamberlin that the earth has always been solid. In the first place, Chamberlin has accepted the opinion of Osmond Fisher that thermal contraction is insufficient to account for mountain building, and in place of it has suggested that accretion would give the requisite compression. Matter deposited on the surface of the earth would compress the interior, and in consequence of this contraction the interior would be too small for the exterior and would produce crumpling. In this way a linear crumpling comparable with the radius of the earth itself could, according to Chamberlin, be produced. It will be seen, however, that this could be true only if the earth's radius had been increased by accretion by a large fraction of itself, which has just been seen to be impossible. Further, the compression would be a slow process. The matter deposited on the surface would necessarily be in a fragmentary condition. It could only begin to be consolidated by pressure when it had been buried to such a depth that the pressure was sufficient to cause plastic deformation. Then gradual flow would commence, accompanied by reduction of volume. Meanwhile the matter above it, being under less pressure, has not begun to be consolidated. Contraction below therefore produces no crumpling; the fragments merely roll and slide over each other. Thus on the planetesimal theory there could at no stage be crumpling at the surface till accretion had ceased. Afterwards crumpling could arise only through internal cooling, which would necessarily be less than was possible in a fluid earth, and through the continuance of the consolidation of the interior. Now plastic flow once started in any region would continue until adjustment was complete, and would for ordinary substances be rapid in comparison with the slow process of accretion. Thus at most a few kilometres in depth would be capable of flow under the actual pressures existing, without having already adjusted themselves completely before the end of accretion.

* *M.N.R.A.S.* 77, 1916, 84–112.

Consolidation could accordingly contribute at most only a few kilometres to the available compression at the surface. The Planetesimal theory is therefore able to explain less mountain building than the fluid earth theory.

A·3. *The History of the Atmosphere.* It has also been argued that a fluid earth could not have retained any atmosphere, and especially water vapour. This argument is capable of two independent replies: first, that the earth could have retained its water vapour, and second, that even if it had lost its whole atmosphere it could have produced a new one. It is seen that Chamberlin's argument supposes that if the earth's surface reached a temperature of 1500° A. most of the gases of the atmosphere would depart from the influence of its gravitation. If C is the mean square velocity in a gas, Jeans has shown that half the gas would be lost in 2×10^6 years if C was as high as 2·5 km./sec. For hydrogen (H_2) at 280° A., C is 1·9 km./sec. Thus hydrogen could certainly be retained at ordinary temperatures for a longer time than has been considered could have elapsed between the formation of the earth and the cooling of its surface. Now C is proportional to $m^{-\frac{1}{2}} V^{\frac{1}{2}}$, where m is the molecular weight and V the absolute temperature. For water vapour at 1500°, C would therefore be only 1·47 km./sec.; thus water vapour could be easily retained for the time required. Still more so oxygen and nitrogen would be retained. If the earth was primitively distended, water vapour could still be retained when the radius was three times as great as at present, since the critical velocity is proportional to the reciprocal of the square root of the radius. There is therefore no insuperable obstacle to the retention of water vapour and all heavier constituents of the atmosphere when the earth was fluid.

A·31. It is possible, however, that very little of the atmosphere is primitive. We know that hydrogen, water vapour, carbon dioxide, and nitrogen are being continually evolved from volcanoes, and it is possible that most of the atmosphere has been produced from this source since solidification. This requires all our water to have been within the earth initially. A few years ago the idea of water at 1500° mixed with siliceous constituents would have seemed ridiculous, but it is now known that at high temperatures and pressures water and rocks actually mix freely in all proportions. Thus it is probable that all the water in the primitive atmosphere was absorbed by the rocks and has only been given off again since solidification*. It was for a long time considered difficult to explain how igneous rocks came to contain water of crystallization, but this problem may now be regarded as solved.

Oxygen is not produced to any appreciable extent by volcanoes, but carbon dioxide is, and it is possible that all the oxygen in the atmosphere has been formed from carbon dioxide by the action of plants. There is a difficulty in this suggestion, though perhaps not an insuperable one. Ordinary carbon assimilation, by means of chlorophyll, is characteristic of the higher plants. The lower plants absorb oxygen and expire carbon dioxide like animals; so do even the green plants when in the dark. Thus it is not easy to see how the conversion of carbon dioxide into oxygen could have commenced until plant life had reached a fairly advanced stage in evolution, and until that stage was reached the plants must have

* J. W. Evans, *Observatory*, **42**, 1919, 165–167.

lacked the oxygen they needed. If this objection is valid the oxygen in the atmosphere, or at any rate some of it, must be primitive. Some plants, such as the nitrifying bacteria, use chemical instead of radiative energy to decompose CO_2, but even they require oxygen initially*. This 'chemosynthesis' suggests a solution, but does not complete it.

Hydrogen is produced in quite sufficient quantities to account for the trace of it present in the atmosphere; most of it combines soon with oxygen. Helium is produced by radioactivity. The difficulty in this case is to understand why so little of it is present in the atmosphere. With the data of Chapter VIII we can show that average acid rock produces helium at the rate of $1 \cdot 7 \times 10^{-12}$ c.c. of helium per gram of rock per year. The helium at the earth's surface is responsible for about 1/250,000 of the atmospheric pressure; thus if all the helium in the atmosphere were under ordinary atmospheric pressure it would make a layer over the surface 3·2 cm. thick. If now 1000 c.c. of acid rock were denuded from each square centimetre of the earth's surface, and all their helium transferred to the atmosphere, they would give 5×10^{-9} c.c. helium per cm.^{2} of the surface for each year of the interval between the formation of the rock and its denudation. If then the average age of igneous rocks when denuded was 160 million years, all the helium of the atmosphere would be explained if the average thickness of igneous rocks denuded from the earth's surface was 40 metres. This is an impossibly small amount. It could be increased if a large fraction of the helium in rocks is retained in the sedimentary rocks formed from them, but it appears unlikely that this fraction exceeds $\frac{1}{2}$, and thus the depth can only be doubled. No satisfactory explanation of why the atmosphere contains so little helium has yet been offered†.

On the theory of the former fluidity of the earth it is therefore possible to account for the gases of the atmosphere. On Chamberlin's theory, however, it is not possible to account for the existence of the atmosphere. The small gravitative power of his small earth nucleus would have made it unable to retain permanent gases during its aggregation, and any gases or water absorbed in it could never have got out again, since it was by hypothesis buried below some thousands of kilometres of planetesimals. The atmosphere must then have been brought by the planetesimals. But the planetesimals, like modern meteorites, must have been completely arid and atmosphereless. Thus, as in the previous case, Chamberlin's arguments are more injurious to his own theory than to the one he is attacking.

A·4. *The Primitive Crust*. It has been suggested that the primitive crust of a formerly solid earth should be in a characteristic condition, and might be geologically recognizable. No trace of the primitive crust has been found, and some geologists have made this fact the basis of an attack on the theory of the former fluidity of the earth. In view, however, of

* Pfeffer, *Physiology of Plants*, 1, 1900, 361–363.

† I believe that the first edition of this book contained the first publication of this point. It should be stated that Holmes alluded to it in a letter to me in 1917, and that Prof. F. A. Lindemann mentioned it at the meeting of the Royal Society in 1921, when Prof. H. N. Russell's paper was read (cf. p. 72). The recent work of Lindemann and Dobson, indicating a rise of temperature with height to at least 300° A., may help to enable the excess of helium to escape and thereby provide a solution.

the extent of denudation and redeposition that we know to have occurred, there would be no cause for surprise if the whole of the primitive crust should have been removed or buried. The argument, so far as it has any validity, is, of course, equally applicable against the planetesimal theory, for planetesimal matter unaltered since deposition would also be in an easily recognizable condition, and has not been identified.

A·5. *The Land and Water Hemispheres.* A further argument has been based on the existence of the land and water hemispheres. One of the outstanding problems of geophysics is the mechanism that produced the marked asymmetry of the distribution of land and sea. Some possible explanations of this have been discussed in 15·7. It has been suggested that the phenomenon is explicable on the planetesimal theory. More planetesimals might have fallen on one side of the earth than on the other, and thus have produced the observed inequality. The probability of such a hypothesis is, however, inappreciable.

Let the radius of the earth be a, and consider the probability of any planetesimal falling in a range of longitude $2\pi\alpha$ of the surface. This probability is evidently α; for on account of the rotation of the earth, all meridians are equally frequently presented to the planetesimals, in whatever direction they may approach. Suppose then that n planetesimals fall altogether. The prior probability that m of them fall within the sector $2\pi\alpha$, the fall of each being unaffected by those that have fallen already, is

$$^nC_m\,(1-\alpha)^{n-m}\,(\alpha)^m \qquad \text{.........(1)},$$

where ^nC_m denotes the number of combinations of n things taken m at a time. Let us denote this probability by P. This is a maximum if $m/n=\alpha$. Thus a uniform distribution in longitude is the most probable. Let us now investigate the probability of a departure from the uniform distribution by a given amount. Put

$$m/n = \alpha + \xi \qquad \text{.........(2)},$$

where ξ is small and $n\xi^2$ moderate. Then we may approximate by Stirling's formula

$$\log n! \sim -n + (n+\tfrac{1}{2})\log n + \tfrac{1}{2}\log 2\pi \qquad \text{.........(3)},$$

and obtain, when n is large and neither α nor $1-\alpha$ very small,

$$\log P = -\tfrac{1}{2}\log 2\pi n\alpha(1-\alpha) - \frac{n\xi^2}{2\alpha(1-\alpha)} \qquad \text{.........(4)}$$

$$+ \text{ terms of order } \xi \text{ and } \xi^2.$$

Now if we consider the range between $n(\alpha+\xi)$ and $n(\alpha+\xi+d\xi)$, where $d\xi$ is small compared with $n^{-\frac{1}{2}}$, the number of possible values of m in this range is clearly $nd\xi$. Thus the probability that m lies in this range is

$$\left\{\frac{n}{2\pi\alpha(1-\alpha)}\right\}^{\frac{1}{2}} e^{-n\xi^2/2\alpha(1-\alpha)}\,d\xi \qquad \text{.........(5)}.$$

The probability that m exceeds $n(\alpha+\xi)$ is therefore

$$Q = \left\{\frac{n}{2\pi\alpha(1-\alpha)}\right\}^{\frac{1}{2}} \int_\xi^\infty e^{-n\xi^2/2\alpha(1-\alpha)}\,d\xi$$

$$= \tfrac{1}{2}\left\{1 - \operatorname{Erf}\xi\left(\frac{n}{2\alpha(1-\alpha)}\right)^{\frac{1}{2}}\right\} \qquad \text{.........(6)}.$$

Now when the argument is great, an approximation to Erf x is given by the formula

$$1 - \operatorname{Erf} x = \frac{e^{-x^2}}{x\sqrt{\pi}} \qquad \text{.........(7)},$$

so that

$$Q = \left(\frac{\alpha(1-\alpha)}{2n\pi}\right)^{\frac{1}{2}} \frac{1}{\xi} e^{-n\xi^2/2\alpha(1-\alpha)} \qquad \text{.........(8)}.$$

In our problem n is the total number of planetesimals picked up by the earth. If then b is the radius of a planetesimal, and if the radius of the earth has increased by accretion by $\frac{1}{4}$ of its original value, we have

$$n = a^3\{1 - (\tfrac{4}{5})^3\}/b^3 = 0{\cdot}5a^3/b^3 \text{ nearly} \qquad \text{.........(9)}.$$

The Pacific Ocean is, on an average, about four kilometres deep, and covers about half the earth. Thus if we consider the probability of getting such a depth by accident, as has been suggested, we take $\alpha = \frac{1}{2}$ and $\xi = \frac{1}{2000}$. Thus

$$Q = \left(\frac{b^3}{4\pi a^3}\right)^{\frac{1}{2}} 2000e^{-a^3/4\times 10^6 b^3} \qquad \text{.........(10)}.$$

Evidently the smaller the planetesimals the smaller is the probability of the hypothesis. We can hardly suppose the planetesimal to have a larger radius than a kilometre, in which case $a/b = 6000$. Then

$$Q = \frac{1}{800} e^{-50,000} \qquad \text{.........(11)},$$

which is utterly insignificant. Thus the formation of the land and water hemispheres in the process of accretion is practically impossible.

Chamberlin suggests in *The Origin of the Earth* that denudation during growth would increase the asymmetry, even though this might be originally insignificant. This does not, however, appear likely. If an ocean formed early on a nearly spherical nucleus, as he suggests, it would submerge any small inequality that might arise, and the only denudation possible would be the insignificant amount that goes on at the ocean bottom. This hypothesis does not therefore appear to solve any difficulty presented by the fluid earth theory.

A·6. I believe, therefore, that the Planetesimal Hypothesis does not offer a solution of any of the great outstanding problems of geophysics, and that on cosmogonical grounds it is quite unacceptable. It has, however, been of considerable value in the past in suggesting new lines of attack on our problems, and the tidal theory here adopted arose as a modification of the Planetesimal Hypothesis designed to avoid the objections that appeared fatal to it in its original form. To reject this hypothesis now is not to deny its importance in the history of cosmogony and geophysics.

APPENDIX B

Jeans's Theory

"It is an old maxim of mine that when you have excluded the impossible, whatever remains, however improbable, must be the truth."
A. CONAN DOYLE, *The Adventures of Sherlock Holmes.*

B·1. The chief point of difference between the theory of Chapters II, III and IV and that of Jeans* is that whereas I suppose the primitive sun at the time of the disruption to have had approximately its present size, he supposed it to have been distended so as to include the orbit of Neptune. There are several serious objections to such a distension of the sun. We cannot assume the sun to have been much more massive than it is now, and the astrophysical evidence against so small a star having such a distension and a temperature consistent with the gaseous state is overwhelming. Also planets produced at distances from the sun greater than the present distance of Neptune would have had to be brought in by some means. The only agent available for this purpose appears to be a resisting medium. It has been shown in 4·2 that a resisting medium, the motion at any point of which was not sensibly different from that of a planet in a circular orbit at the same point, could have no effect on the mean distance of a planet; and that a medium for which this difference was considerable would have too small a density at such a distance to produce any appreciable frictional effect on a planet. Thus in neither case could the mean distances be affected appreciably by a resisting medium.

B·2. *Prior Probability of a Disruptive Encounter.* Jeans's reason for adopting such a distension was based on the belief that the prior probability of any encounter close enough to break up the sun when less distended is so small as to forbid it. This, however, does not seem to be the case. Jeans estimated in 1919 that in the present universe the average interval between encounters at a distance less than $1{\cdot}3 \times 10^{15}$ cm. is about 3×10^{10} years. Thus the probability that in any one year any particular star that we have no other data about will have an encounter at this distance or less is $1/(3 \times 10^{10})$. This probability is proportional to the square of the distance of approach; therefore the probability that any one star would have one at a distance of 10^{11} cm. is $1/(5 \times 10^{18})$. Hence the probability that any particular star would have an encounter within the latter distance in 10^{11} years, a time probably short compared with the age of the sun, is $1 - \{1 - 1/(5 \times 10^{18})\}^{10^{11}} = 1/(5 \times 10^{7})$, and the probability that one star in the whole universe, containing perhaps 10^{9} stars, would have an encounter at this distance, is $1 - \{1 - 1/(5 \times 10^{7})\}^{10^{9}}$ or practically unity. Thus it is practically certain that at least one star in the universe has had an encounter at the distance here assumed, and our only datum in the matter is that there is at least one solar system. Thus

* The opinion of Jeans referred to here was expressed in his *Problems of Cosmogony*, 1919. He has since accepted my view (*Astronomy and Cosmogony*, 1928), but his earlier arguments are worth discussing irrespective of whether he is still convinced by them.

the hypothesis that the diameter of the sun at the encounter was comparable with its present one is consistent with knowledge of stellar motions. It is, however, probable that systems of the type of the solar system are the exception in the universe and not the rule.

B·21. It may be remarked that the tidal theory of the origin of the solar system might have a considerable probability on the evidence before us, even if stellar encounters were much less frequent than has been inferred from the present condition of the stellar system. For, in the notation used by Dr Wrinch and myself*, let h denote the aggregate of propositions believed independently of experience, and let $P(p:qh)$ denote the probability that a proposition p is true, given that h and another proposition or set of propositions q are true. Let $\sim p$ denote the proposition that p is untrue, and pq the proposition that p and q are both true. Let the empirical data of physics, not including the existence of the solar system, be denoted by k, and let the proposition that the solar system exists be denoted by m. The laws of physics, which we may denote by l, have a high probability on the data k, so that

$$P(l:kh) = 1 - \alpha \qquad \text{.........(1)},$$

where α is very small.

The arguments for the tidal theory fall into two groups. First, it is shown that, given our physical laws among the data, no theory other than a tidal one can account for the solar system as it is. If then t denotes the proposition that the initial conditions required by the tidal theory once existed, this result may be expressed by

$$P(m:\sim t.lkh) = \beta \qquad \text{.........(2)},$$

where β is very small. It would be zero if the proof possessed strict mathematical accuracy, and the proof available is enough to show that β must be exceedingly small.

Next, it is shown that the tidal theory is capable of explaining the main features of our system, given suitable initial conditions. Thus

$$P(m:tlkh) = \gamma \qquad \text{.........(3)},$$

where γ is certainly not small and may be nearly unity.

Now we have the general propositions of the theory of probability

$$P(pq:h) + P(p.\sim q:h) = P(p:h) \qquad \text{.........(4)},$$

$$P(pq:h) = P(p:h)\,P(q:ph) \qquad \text{.........(5)}.$$

Then

$$\begin{aligned} P(m:kh) &= P(mlt:kh) + P(ml.\sim t:kh) + P(m.\sim l:kh) \\ &= P(l:kh)\,P(t:lkh)\,P(m:tlkh) \\ &\quad + P(l:kh)\,P(\sim t:lkh)\,P(m:\sim t.lkh) \\ &\quad + P(\sim l:kh)\,P(m:\sim l.kh) \qquad \text{.........(6)}. \end{aligned}$$

Let us write

$$P(t:lkh) = \tau,$$

$$P(m:\sim l.kh) = \delta.$$

Then (6) becomes

$$P(m:kh) = (1-\alpha)\{\tau\gamma + (1-\tau)\beta\} + \alpha\delta \qquad \text{.........(7)}.$$

Again,

$$P(mlt:kh) = (1-\alpha)\,\tau\gamma,$$

* *Phil. Mag.* **38**, 1919, 717–731.

being the first term in (6), and also

$$= P\,(m : kh)\, P\,(lt : mkh) \qquad \text{.........(8).}$$

Combining (7) and (8), we have

$$P\,(lt : mkh) = \frac{(1-\alpha)\,\tau\gamma}{(1-\alpha)\,\{\tau\gamma + (1-\tau)\,\beta\} + \alpha\delta} \qquad \text{.........(9).}$$

Now δ, being a probability number, is necessarily not greater than 1, and α and β are exceedingly small. Thus the expression on the right will be practically unity unless $\tau\gamma$ is so small as to be comparable with either α or β. Omitting this alternative for a moment, we see that $P\,(lt : mkh)$ is practically unity. In other words, given the empirical data of physics and the existence of the solar system as it is, it is practically certain both that the laws of physics used in the argument are true and that the initial conditions required by the tidal theory once occurred; and therefore that the tidal theory is true.

This argument would break down if $\tau\gamma$ was very small. We have seen in B·2 that τ is probably nearly unity, and that γ is moderate, and therefore there is no question of its failure with the data here used. For failure it would be necessary for $\tau\gamma$ to be less than one of the two quantities α and β. In other words, the probability of there having ever been, during the life of a single star in the universe, an encounter at a distance less than 10^{11} cm., must be less than either the probability of a fallacy in the argument against the possibility of the formation of the solar system by purely internal action, or the probability that the laws of physics are wrong. This gives a conception of the strength of the position of the tidal theory. If β is the greatest of the quantities $\tau\gamma$, α and β, $P\,(m : kh)$ will be practically

$$(1-\alpha)\,(1-\tau)\,\beta = P\,(ml\,.\!\sim t : kh)$$
$$= P\,(m : kh)\,P\,(l\,.\!\sim t : mkh),$$

and therefore $P\,(l\,.\!\sim t : mkh)$ will be practically unity. Thus it will be practically certain that the solar system was formed by some non-tidal action and that the laws of motion are right. If, again, α is the greatest, it will follow that the laws of motion are probably wrong. Thus we see that the extreme smallness of $\gamma\tau$, besides forcing us to reject the tidal theory, would make other much more bizarre inferences necessary.

APPENDIX C

The Relation of Mathematical Physics to Geology

"If you make some compliments to a man about that which he knows he does well, that is no good. But if you praise him for what he do pretty dambad, then you give him a great pleasure. Suppose you go to a painter and say to him: 'Your painting is exquisite. Your pictures are of a sublimity.' Well, you do yourself no good. He laugh and change the subject. But if you go to that same painter and say: 'But what a ravishing voice! Why are you not in Grand Opera?' ...then you may perhaps borrow half-a-crown till next Thursday."

BARRY PAIN, *Confessions of Alphonse.*

C·1. Geophysics is by nature a subject that draws its data from many widely different sources: pure geology, astronomy, seismology, geodesy, and the investigation of the earth's thermal state being among the chief. Its task is to coordinate the information provided by them, and its methods are those of theoretical physics.

It is often found that the results of one method of inquiry afford useful information in another field; and so far as I know there is no case where there is an unequivocal contradiction between two sets of data relevant to the same point. If we believe in a principle of causality we shall expect this; if we are merely hopeful of the utility of scientific method we shall welcome it. But investigators are liable to have very different opinions as to where their data end and where inference from them begins, and it often happens that the superstructure erected on one set of data contains a hypothetical portion capable of being adequately tested only by information collected in an entirely different field. The development of geophysics and of the associated sciences is therefore essentially a matter of cooperation between workers whose methods are different; and it is essentially difficult on account of their differences in training and resulting outlook. Nevertheless the progress of science demands that it should be attempted.

It must be confessed that this necessity is far from being adequately realized. A widespread attitude is expressed in the following remarks, which are slightly condensed from those of a well-known geologist. "It must be said that Wegener is not convincing as to the earth forces available for the production of these mighty mass movements. Jeffreys, indeed, has stated that the known forces tending to move the continents are 10^{14} times too small! But geologists will have Huxley's mathematical mill too much in mind to pay much heed to this latest physico-mathematical opinion on a geological problem; and if adequate geological proof of continental drift is forthcoming, geologists will believe it despite any assertion of its physical or mathematical impossibility." I do not in the least believe that the author of this very clear statement intended to order anybody that knows any mathematics or physics to get off the earth; but nevertheless his remarks amount to a denial of the right of geophysics to exist.

Since Huxley's dictum on the mathematical mill is usually quoted in this connexion, it may be well to give it in full*. It concerned the earlier

* *Q.J.G.S.* **25**, 1869, *l.*

discussions on the duration of geological time. "I do not presume to throw the slightest doubt upon the accuracy of any of the calculations made by such distinguished mathematicians as those who have made the suggestions I have cited. On the contrary, it is necessary to my argument to assume that they are all correct. But I desire to point out that this seems to be one of the many cases in which the admitted accuracy of mathematical processes is allowed to throw a wholly inadmissible appearance of authority over the results obtained by them. Mathematics may be compared to a mill of exquisite workmanship, which grinds you stuff of any degree of fineness; but, nevertheless, what you get out depends on what you put in; and as the grandest mill in the world will not extract wheat flour from peascods, so pages of formulae will not get a definite result out of loose data."

Now it may be surprising to some geologists (I must expressly say not all) to be told that there is nothing in the above passage that is not so familiar to every mathematical physicist as to need no repetition. As to the special topic under discussion, it is now recognized by physicists and geologists in general that both sides were wrong at the time: the physicists because they were then unaware of an essential factor in the problem (radioactivity), and the geologists because they regarded as a measure the denudational method, which could at best give only an order of magnitude. The former was an error of ignorance, the latter one of principle; and it is by a combination of physical and geological methods that the problem has at last been solved. The history of this problem should be sufficient indication of the need for cooperation.

Now as to the appearance of accuracy given by mathematical methods, everyone using them is aware that inaccuracies of the data are reflected in the results. Often they are casual errors of observation and can be expressed by means of a probable error; but in geophysical problems there are often also systematic errors due to factors not yet evaluated. These can very often be dealt with by considering extreme cases, between which the truth must lie; and for a great many purposes this is adequate. Often, again, the discrepancy when a doubtful hypothesis is under discussion is such that no reasonable modification of the data offers the slightest prospect of producing an agreement. In such a case it seems legitimate to give the result as it stands for the advocate of such a hypothesis to make any modification he thinks fit. But in any case I would point out that uncertainties in data that will bear a moment's inspection afford no excuse for not working out their consequences. The mathematical mill works in both directions. We may not be sure of what we are putting in; but if what comes out is an indigestible form of cellulose, it is legitimate to infer that what was put in was not wheat; and disagreement of inferred results with the facts provides one of the best means of correcting our data and increasing our knowledge. Again, it is sometimes found that the results of one line of inquiry are in definite disagreement with the alleged facts of another. When this occurs it is always found that an element of hypothesis has entered somewhere without being noticed; and it is in just this way that such tacit hypotheses can be detected. Unfortunately, however, there is nothing that many investigators dislike more than having their unstated assumptions pointed out; if this were not so, scientific controversy would be a much more amicable and profitable thing than it often is.

C·2. A curious opinion prevalent among geologists is that geology deals with facts, geophysics (which they are liable to call mathematics) with *a priori* considerations. For instance, many geologists have been kindly patronizing to me on the subject of the identification of the intermediate layer with tachylyte. Now I cannot claim the credit for inventing the idea that there is basaltic matter within the earth in a glassy and not a crystalline state; it is due to a distinguished geologist, Prof. Daly. My innovation was to recognize that two sets of facts discovered after Daly's suggestion, namely the velocities of the P^* and P waves, and the laboratory determinations of the compressibilities of crystalline basalt and tachylyte, implied directly that tachylyte, if present at all, forms the intermediate layer and not the lower one. The objections of other geologists to a tachylyte layer are based on a belief that magma solidifying at a considerable depth would take a coarsely crystalline or gabbroid form. But this is a purely *a priori* consideration. No geologist has obtained a specimen of the matter *in situ* 20 km. down, and the question at issue is whether this matter has ever crystallized at all; there is at present no direct evidence as to the method of solidification of a basaltic magma when it contains volatile constituents in such quantity as to lower its melting point to 700° at most, and the escape of these constituents is prevented by high pressure. The question of the presence of gabbro, on the other hand, is one of fact, for the seismological and laboratory evidence taken together imply that there is no widespread layer of this rock. The only alternative that seems worth considering at present to the opinion that basalts observed at the surface have come from an intermediate layer of tachylyte is that they have come from a layer of eclogite; gabbro is out of the question in any case. The absence of laboratory measures of the elasticity of eclogite make all allusions to it speculative, but even should they confirm the possibility that either the lower or the intermediate layer is eclogite it will be necessary to take into consideration the fact that the garnets decompose at temperatures below their melting points: thus Boeke and Eitel remark (p. 290) that "the conditions of formation of the garnets so far remain in complete darkness."

I do not wish to imply that I consider the identification of the intermediate layer with tachylyte and the lower one with dunite as certainly right, or Holmes's diorite-eclogite-dunite succession as impossible. I regard them both as working hypotheses capable of further test. My reasons for preferring the former are simply that it is known to be in accordance with the seismological facts, while the other has not yet been compared with them at an essential point; and that it also fits the thermal data without our having to assume that the materials in depth have markedly different radioactive contents from materials of similar composition at the surface. New evidence may alter this situation, but meanwhile we must accept it.

C·3. The hypothesis of continental drift has recently obtained a great deal of attention. Criticism of this hypothesis falls naturally under three heads. First, Wegener has suggested several mechanical forces tending to produce continental drift: are these adequate for the explanation of the drift asserted to have taken place? This question has already been answered in the negative in 15·8. Second, some of his supporters suggest that there may be other forces not considered that could produce such an effect. One may query in return whether, if such a force were found,

it would remain the same theory; or are such theories like the antique chair that remained an antique chair after it had had a new back, a new seat, and three new legs fitted to it? But as a matter of fact some approach to a definite answer to this view can be made. In a highly viscous material such as constitutes most of the earth permanent shear, if any, is in the direction of the force. (This is not true in the atmosphere or the ocean; it is a question of the relative sizes of the rotational and viscous terms in the equations of horizontal motion.) To make America move westwards with respect to the Old World, as is required by the theory, a westerly force is required; the only one known is tidal friction, which is on an average under 10^{-7} of the equatorial drift, and therefore at the most favourable estimate would take 10^{17} years to produce the desired effect. To produce the effect in 3×10^7 years we require a westerly force 10^{10} times as potent as tidal friction. But tidal friction is not an insignificant force; it has altered the earth's rate of rotation very considerably during its history, and probably during geological time. It appears that such an external force as the modified theory would require would stop the earth's rotation in a time of the order of a year.

The third type of criticism is the geological one: accepting Wegener's reconstruction of the primitive continental mass, several other implications arise, of a purely geological nature*. Some of these are favourable to the reconstruction; others are definitely against it, but in many discussions only the former are mentioned. For instance, the alleged fit of South America into the angle of Africa is seen on a moment's examination of a globe to be really a misfit of about 15°. The coasts along the arms of the angle could not be brought within several hundreds of kilometres of each other without distortion. The widths of the shallow margins of the ocean near the continents lend no support to the idea that the forms have been altered considerably by denudation and redeposition; and if the forms had been altered by folding there would be great mountain ranges at a distance from the angles with their axes pointing towards the angles, which is not the case. (The Brazilian Heights are greatest *near* the angle, where the distortion required is least.) Similar misfits are encountered in comparing North America with Europe†. A petrological comparison of the regions alleged to have been in contact leads to a further set of inconsistencies‡. Until these geological and geographical objections are answered it hardly seems worth while discussing the physics of the theory any further. But reference must be made to the assertion that the situation of the Rocky Mountains and the Andes is what would be expected on the theory of continental drift. On the contrary, it is one of the most definite pieces of evidence against it. Either the materials of the ocean floor are stronger than those of the continents, or they are weaker. If they are stronger they will not give way to let the continents move through them; if they are weaker, the continents would advance, if at all, without being fractured, and no mountains would be formed. Considering how often it has been said that Wegener's theory explains the Pacific mountains, it is

* Cf. the reactions of Dr Rastall, Mr Lake, Prof. Holmes, and Prof. Schuchert to recent works by Du Toit and van der Gracht: *Geol. Mag.* 1928, 139–140, 422–424; *Nature*, **122**, 1928, 431–433; *Am. J. Sci.* **16**, 1928, 266–274.

† P. Lake, *Geol. Mag.* **59**, 1922, 338–346; *Geog. Journ.* **61**, 1923, 179–187.

‡ H. S. Washington, *J. Wash. Acad. Sci.* **13**, 1923, 339–347.

odd that my explanation of them, first given in 1921, has hitherto received no mention in geological literature*.

C·4. The theory of Joly raises much deeper issues. It is based on the assumption that the radioactive matter in the earth's crust is sufficient and suitably distributed to make the ultimate steady temperature S_0 of Chapter VIII above the melting points of the deeper-seated rocks. The difficulty that this implies permanent fusion at all depths below some tens of kilometres, which is decisively contradicted by every available type of evidence, is met by supposing that the earth's present state is not representative. The deep-seated matter is supposed to melt at intervals, giving epochs of great igneous activity; tidal friction then tows the outer crust round over the fused interior until the excess of heat has been conducted through the ocean floor, and the whole solidifies again. On this theory we are now in a stage such that the heat being conducted out of the earth is less than that being generated within it.

Now such a course of events is contrary to both theory and experiment in every case of heat transfer that can be conceived as showing the slightest resemblance to the conditions within the earth's crust. In problems of heat conduction in solids the universal experience when there is a steady internal source of heat with free radiation from the surface is that the temperature at every point tends asymptotically to a steady value. Alternation with time, such as is the normal event in problems of elasticity, does not occur in heat conduction, on account of the fundamental difference that the time enters problems of the former type through $\partial^2/\partial t^2$, the latter through $\partial/\partial t$. Conditions are similar, and the same theory is applicable, in the liquid state when stable. When a liquid is heated below, however, it usually becomes unstable and convection currents are set up. These redistribute the heat much more rapidly and thoroughly than simple conduction can. The result is equivalent to a great increase in the thermal conductivity, so great that a liquid in a convective state can usually be regarded as infinitely conducting in comparison with a solid. Latent heat of change of state expresses a purely conservative tendency; a solid actually at the melting point does not rise in temperature if the heat generated internally exceeds that conducted out, and a liquid at the melting point does not cool in the opposite situation; there is nothing in this making for alternation. The presumption is, therefore, that if the distribution of radioactive matter is such that the calculated ultimate basal temperature is above the melting point, the system will tend to a state where there is a permanently fluid layer; spontaneous resolidification and subsequent repetition of the process are out of the question.

Prof. Joly complicates the question by assuming that a complete layer of the crust, beneath continents and oceans alike, fuses at one time, and that tidal friction then makes the crust revolve over the fused layer†. This does not remove the difficulty, because the tendency to asymptotic approach to a steady state would only make the depth of fluid below a given part of the surface alternate within a range whose extremes are both intermediate between the depths appropriate to oceanic and continental regions in the steady state. The depth would vary, but could never vanish, and there would be no resolidification. This hypothesis is open to the further

* *Proc. Roy. Soc.* A, **100**, 1921, 127–128.

† *Phil. Mag.* **45**, 1923, 1167–1188; *The Surface History of the Earth*, 1925.

objection on its own account that however we choose the depth of the layer and its viscosity we cannot make tidal friction produce the required effects in the time postulated by Prof. Joly.

These objections have been published previously in greater detail, and Prof. Joly's replies have consisted of a series of evasions and irrelevancies*. But I must repeat that my main objection to his theory is not that there are serious arguments against it; it is that there are no serious arguments for it. It is not the task of the opponent of a theory to examine every possible variation of it to make sure that there is none that will give the kind of effects that it is claimed to explain; it is the task of the proposer to produce one form of the theory that will give these effects, to state it in a form sufficiently precise for its consequences to be worked out, to investigate these consequences by the known methods of theoretical physics, and to compare them with the facts. When this is done there is a basis for discussion. But when, as in this case, the hypotheses as to facts are vaguely stated, when the physical laws assumed as to transfer of heat, rate of melting, and the mechanical phenomena concerned are not stated at all, and when no attempt is made to investigate their consequences properly, all it is possible to do is to point out the obvious difficulties and to leave the author to overcome them if he can. In this case he has not thought it necessary.

C·5. These principles, it may be said, are very generally recognized. They are the rules that determine the acceptance or otherwise of papers for publication by the chief mathematical and physical societies, and I have no reason to doubt that they are also applied in other subjects. But Prof. Joly, who is an editor of the *Philosophical Magazine*, and first published his theory in that Journal, declares that they are of no interest, and sees in my enunciation of them nothing but an expression of personal irritation.

If anything it is in a border-line subject such as geophysics more than anywhere else that a due sense of responsibility in scientific publication is necessary. It is inevitable in such a subject that a new theory making far-reaching claims will attract much attention among workers whose training has not been such that they can detect its weak points on inspection; in a more specialized subject, such as petrology, electron theory, or tidal dynamics, this is not so. Thus I have seen Prof. Joly's theory described as a 'simple and convincing exposition'! The result is a curious distribution of publicity, depending not on what a theory has explained, but on what its author claims to have explained. I find, for instance, that for once that I am asked for my views on any subject where I have constructive remarks to make, I am asked ten times for details of my objections to the theories of Joly and Wegener.

On the other hand a theory that coordinates a large number of facts, but is sufficiently specific to be understood and does not claim to have explained everything, seems to have very great difficulty in getting even its successes recognized. Thus a well-known text-book published last year dismissed the thermal contraction theory of mountain formation with the remark that in the opinion of most geologists the amount of contraction accounted for was inadequate. Seeing that it has been known for twelve years that the amount is of the correct order of magnitude, and that no

* *Phil. Mag.* 1, 1926, 932–939; 4, 1927, 338–348; 5, 1928, 215–221.

data are available for an exact comparison, such a remark is singularly unfortunate. Holmes, who is the only critic of the theory that appears to have read any work on it under twenty years old, does not base his objections on quantitative inadequacy. But the least happy feature of the remark is its suggestion that the ultimate test in science as in politics is in the number of votes secured, and not in the evidence and arguments available.

I have not thought it necessary to give references to the extracts from recent geological writings quoted above, because they are representative of a wide class of opinion; they have been chosen only because they have as a matter of fact been expressed in a convenient form for quotation. I am fully aware that the views are not universal among geologists, though they are common; in fact my indebtedness to geologists that do realize the need for continual cooperation between their own methods and those of physics is obvious throughout this work. May I close with the hope that the latter class may increase and absorb the former?

APPENDIX D

Theories of Climatic Variation

"Now is the winter of our discontent
Made glorious summer."

SHAKESPEARE, *King Richard III*, Act I, Sc. 1.

D·1. *The Geological Evidence for Climatic Changes.* Climatic variation as such is essentially a meteorological topic, and therefore is outside the scope of the present work. On the other hand, climatic variations have produced very great effects on the surface features of the earth in the past, and their causes may be largely drawn from the non-meteorological parts of geophysics. Since they serve as a connecting link between two problems definitely within our scope, some discussion of them may be in place.

There is a great deal of evidence for considerable changes in climate in past geological ages. The last glacial period, during which most of Northern Europe and America were buried under a thick ice sheet, is the best known of these vicissitudes; but a glacial period with many similar features occurred in the Pre-Cambrian era, before the oldest known fossiliferous rocks were laid down, and another at the beginning of the Permian period, about 250 million years ago*, besides occasional local glaciations at other times. Between these glaciations there were mild intervals, such as the long spell of warmth in the Secondary and early Tertiary. At present the climates over most of the earth appear to be becoming warmer and drier, though there are places where this does not hold.

D·2. *Suggested Explanations.* Numerous attempts have been made to explain such facts. Most of the hypotheses offered may be classified under the following heads:

A. Changes in the motions of the earth as a whole, especially in
   (1) the eccentricity of its orbit;
   (2) the inclination of its axis of rotation to the ecliptic.

B. Changes in the composition of the atmosphere and ocean, especially
   (1) the amount of carbon dioxide;
   (2) the amount of volcanic dust;
   (3) the amount of salt in the sea.

C. Changes in topography
   (1) in the area of the continents;
   (2) in the height of the continents;
   (3) in the distribution of the continents with regard to one another and to the polar axis.

D. Changes in the sun's radiation.

E. Changes in the internal heat of the earth.

F. The passage of the solar system through cold regions of space.

* Cf. the striking pictures of a Carboniferous boulder-clay and a striated pavement in Daly's *Our Mobile Earth*, 285, 286.

D·3. Of these, A, E, and F may be rejected without difficulty. First, changes in the eccentricity formed the basis of Croll's famous theory, which required that glaciation should occur when the eccentricity was a maximum, and should alternate between the northern and southern hemispheres in the period of the precession of the equinoxes, the glaciated hemisphere being the one containing the pole presented to the sun at perihelion. Both the variation of the eccentricity and the precession of the equinoxes are purely dynamical in origin and extremely regular. Thus glacial periods should have alternated at intervals of less than a million years throughout geological time, which is very far from the state of affairs outlined in the second paragraph above. Changes of eccentricity, however, are real, and may play a subsidiary part. Changes in the inclination of the axis may arise in two ways, first, by planetary perturbations of the plane of the earth's orbit, and second, through internal deformations of the earth itself. The first type of change is regular, and therefore fails for the same reason as Croll's hypothesis; while Sir G. H. Darwin showed that changes in the inclination through internal deformation can never be considerable. The inclination may have undergone some change through tidal friction, but such change would always be in the same direction, whereas the climatic changes have been oscillatory.

D·31. Hypothesis E fails, because, as shown in Chapter VIII, the surface temperature of the earth must have been almost wholly maintained by solar radiation practically ever since it became solid at the surface, and certainly throughout geological time. Conduction from the interior is in comparison quite unimportant.

D·32. Strictly speaking, Hypothesis F is meaningless; there is no such thing as the temperature of space, for temperature is essentially a property of matter. It might, however, be interpreted to mean the temperature that a given body would take up if it was exposed to the radiation in the neighbourhood considered; so that it would be the temperature such that the given body radiated as much heat as it received. But at present the radiation received from other sources is a very small fraction of that received from the sun, and therefore no possible reduction in it could make any appreciable difference to terrestrial temperatures.

D·4. The phenomena enumerated under B must all have occurred to some extent, but it is unlikely that any of them is among the main causes of climatic variation. It has been suggested by W. J. Humphreys* that great volcanic eruptions, such as those of Krakatoa (1883) and Katmai (1912), may produce a lowering of temperature over the earth as a whole for some months; thus volcanic activity on a still larger scale might produce a glacial period. It does not appear, however, that volcanic activity and glaciation have been closely correlated in past epochs. If they had been, great glaciations would have occurred during the times of greatest activity in the Ordovician, Devonian, and Tertiary periods, which is not the case. The absorption of dark heat radiation by carbon dioxide may be appreciable, but it appears probable that the conditions have always been such that any radiation absorbed by CO_2 would have been equally effectively absorbed by water vapour in the absence of the former, so that the thermal effect of variations in the amount of CO_2 in the atmo-

* *Journ. Franklin Inst.* **176**, 1913, 131.

sphere can hardly be great. Variations in the salt content of the ocean have been considered important by Chamberlin. Variation of the density of the ocean from place to place must have some effect in producing a general circulation of the ocean. Thermal expansion near the equator reduces the density of the water, and must therefore have such an effect. If, however, the ocean were much more salt, evaporation of water near the equator would raise the density more than thermal expansion reduces it, and the density circulation would be reversed. This is considered capable of producing an important effect on climate, since the atmosphere is largely heated by contact with the sea. The difficulty in this theory is that the surface waters do not move with the deeper parts of the ocean, but are driven along by the wind*; with the same wind, their motion is practically independent of the density distribution. It is the surface waters alone that affect the temperature of the air. Thus the system of the surface waters and the air is practically self-contained, and the temperature of the air cannot be affected by the salinity of the sea.

D·5. Phenomena of Group C must contribute considerably to the variation of climate. It is known from meteorology that changes in topography could have an important effect on climate, and from geology that such changes have actually taken place; what is uncertain is, first, how far known changes in topography can account for the changes of climate indicated by the geological record, and, second, in cases where prior knowledge of the putative causes is so small that we have to make *ad hoc* hypotheses concerning their time and extent, how numerous are the observed climatic variations that agree with those inferred from the hypotheses.

D·51. The chief work on these lines has been done by C. E. P. Brooks†. His method is based on the correlations between the normal temperatures for January, July, and the whole year, on the one hand, and the nature of the surroundings on the other, found for a network of places over the whole earth. The places he considers are regularly distributed at equal intervals of latitude and longitude all over the earth. Around each he draws a circle of radius 10° of latitude, or 600 nautical miles. The percentages of land and ice in the semicircles to the west and east of the station are recorded separately, and the correlations between these and the temperatures are calculated. From these correlations it is possible to make a quantitative estimate of the effect on temperature both of continentality and of the direction of the prevailing winds. The method is somewhat crude, but certainly takes into account the main features of the problem. The main results are what might be expected from our general knowledge, that land in the neighbourhood of a station makes for extremes of temperature, that ice produces lower temperatures than unglaciated land, and that land or ice produces more effect when it is to windward than when it is to leeward. The special feature of Brooks's work is that it is quantitative as well as qualitative, and can therefore be applied to find the distributions of temperature for other distributions of land and sea. The temperature distributions for other geological dates than the present can therefore be found. Brooks uses temperatures reduced to sea level, so that the effect of altitude must be added before the actual temperature at a station is determined.

* Jeffreys, *Phil. Mag.* **39**, 1920, 578–586.

† *The Evolution of Climate*, Benn Bros. 1922; *Climate through the Ages*, Benn Bros. 1926.

In this way Brooks is able to show that more oceanic conditions, which actually existed, are able to account quantitatively for the mild climate of the Eocene period. A general elevation of the land proceeded throughout the Tertiary Era, and when the Scandinavian highlands and the Rocky Mountains reached the snow line ice sheets commenced to form. Ice becomes colder than unglaciated land exposed so that it receives the same radiation, and produces a greater cooling effect in its neighbourhood. Wind blowing over an unglaciated mountain range reaches the same level on the other side without having been cooled much, or may even have been somewhat heated by the Föhn effect; but wind blowing over an ice sheet is very much cooled when it gets to its original level again. Thus whereas the rising mountains did not produce much disturbance of temperature to leeward before they reached the snow-line, they would afterwards produce a substantial depression. Precipitation upon them would fall as snow instead of rain. That falling on the windward side would form glaciers, and return quickly to the sea; but that to leeward would have no easy outlet, and would accumulate. Thus the ice sheet would tend to spread, especially to leeward. The increased precipitation consequent on the reduction of the temperature of the land would play an important part in the thickening of the ice sheet. The sheet would not necessarily have its highest point above the original mountains; the precipitation to leeward might well raise the surface of the ice to above the original mountain tops, and then the highest point would proceed to move steadily to leeward. The actual events during the glacial period and afterwards agree closely with Brooks's inferences. In particular, the direction of the tips of sand dunes formed at this time gives the direction of the prevailing wind, which corresponds to the effect of a permanent area of high pressure in Scandinavia, and therefore east winds over Germany*. In many other parts of the world striking agreements are found.

A further remarkable contribution was made by F. Kerner-Marilaun†, who introduced a hypothetical 'non-glacial' temperature, designed to sum up the effects of external factors and topography, apart from those of ice. It is the temperature that would be maintained if the ocean had all its actual physical properties, except that the freezing point is supposed sufficiently low for it never to freeze, so that the effect of ice does not arise. This can be calculated more directly than the actual temperature, and forms an intermediate stage in finding the latter. Now it is found that the effect of topography on the mean non-glacial temperature in a given latitude is a few degrees, not enough to account directly for the changes of temperature between a warm period and a glacial one, which is of the order of 30° F. But it appears that the present non-glacial temperature at the North Pole in January is about 26° F. The freezing point of sea water is about 28° F. Thus a rise of temperature of 2° would clear the Arctic Ocean of ice completely, and abolish the effects of floating ice and cold currents, which dominate the climate of so much of the Northern Hemisphere. The long spells of mild climate, which, as Brooks points out, constituted most of geological time, are thus easily explained. The presence of polar ice indicates a glacial period; we are now in a glacial

* Cf. I. Högbom, *Geografiska Annalen*, 5, 1923, 113—142.

† *Sitzber. Akad. Wiss. Wien*, 131, 1922, 153; cf. also Brooks, *Q.J.R. Met. Soc.* 51, 1925, 83-94.

period, though not at its height. Brooks, in the paper just quoted, attempts to evaluate the cooling effect of a polar ice cap, and shows that with a non-glacial temperature at the pole only a few degrees below the actual freezing point, the formation of ice and the consequent secondary cooling would produce glaciation down to latitude 65° and an ultimate depression of the temperature at the pole by 45° F. Thus a slight change in the topography and the external factors affecting the temperatures in high latitudes may make all the difference between a mild period and an intense glaciation.

The whole theory of climatic variation due to topographical changes has been attacked recently by Dr G. C. Simpson*, who claims that the utmost variation of temperature explicable in this way is only a few degrees. But as he does not consider the secondary effect of ice his paper does not deal with the present situation, but with the one that appeared to exist before Brooks's work was done. In his reply to the discussion following the reading of the paper he objected to some of Brooks's physical arguments, but these seem unnecessary to the general validity of the theory. If we grant with Brooks that the lowering of temperature by ice is proportional to the area of ice within a ten-degree circle about the place considered, and determine the constant of proportionality empirically by reference to present conditions, Brooks's results follow. Their physical explanation is a further problem; as a matter of fact I am not in entire agreement with Brooks on the details†, but that has nothing to do with the general question of the effect of ice on surface temperatures and the approximate evaluation of its amount.

Brooks's theory is a very substantial contribution to our understanding of climatic change; but it does not furnish a complete explanation. Glaciation did not begin in many places until long after the mountain tops were not only above the snow line, but above the height where there was any considerable precipitation at all; this is shown by the presence of unglaciated mountain tops, called 'nunataks,' in Scandinavia and the Rockies. It appears as if the later stages, at least, of the elevation of the mountains took place under conditions when the snowfall was inappreciable, and that the ice sheet did not begin to form until some further change of climate, not attributable to the mountains, had supervened. The Caledonian folds, again, must have raised mountains quite comparable with those of the Tertiary, but do not appear to have been followed by glaciation on anything like the same scale, again suggesting that mountain formation, though it may be a necessary preliminary to glaciation, is not a sufficient condition for it. The crustal movements in the last thousand years, again, do not appear enough to account for the climatic changes during that interval.

D·6. Changes in the situation of the polar axis with regard to the land have often been summoned to explain climatic changes that we have no satisfactory explanation for. In particular, Wegener has attributed the Permian glaciation in several parts of the Southern hemisphere to this cause. He supposes that at that time South America, South Africa, India and Australia were all united, meeting somewhere in the South Indian Ocean, and that they have since drifted apart. The presence of a glacial flora in all these places at that time is then explained by the *ad hoc*

* *Q.J.R. Met. Soc.* **53**, 1927, 213–232. † *Met. Mag.* **61**, 1927, 277–281.

hypothesis that the south pole was near the junction. The argument of 15·8 and C·3 seems in itself fatal to this theory. Further, it has been pointed out by Lake* that a similar glaciation took place at this time in Northern Baluchistan, which would on Wegener's hypothesis have been practically on the equator; while Brooks calls attention to the curious distribution of climate in North America at the time, which is inexplicable on Wegener's views, but is reconcilable with the earlier geological ideas of land connexions.

D·7. We now come to the effects of changes in the sun's radiation. We have no means, apart from terrestrial climate, of knowing how the sun's radiation may have changed during geological time, and accordingly any hypothesis of this type is very difficult to test, and therefore very difficult either to prove or to disprove. If any changes were due to this cause, the effects all over the earth should be such as can be attributed to the same change of solar radiation, and this might be thought to afford a test; but unfortunately we cannot determine the dates of events at widely separated places so accurately that we can be sure that the climatic variations at all of them actually took place at the same time; and further, it is far from certain that an increase in the intensity of solar radiation would affect temperatures all over the world in the same direction. W. Köppen† found that the temperature of the earth is highest about sunspot minimum, when the sun is radiating least. This paradoxical result received a partial explanation from H. F. Blanford‡. His suggestion is that high radiation increases the evaporation over the ocean, and therefore increases the cloudiness and rainfall over the land. On both the latter grounds the temperature of the land is reduced. It is to be observed that the temperature of the sea must be raised, for if it were lowered the evaporation would be less than normal, and the argument would have cut off its own feet. If the theory is correct, it can refer only to a limited range of conditions, for it is plain that if the sun's radiation were zero the temperature of the earth would be practically absolute zero, and if it were great enough the oceans would boil and the temperature of the land would be raised to boiling point. Sir G. T. Walker§, on the other hand, has suggested that the effect of sunspots is an increase of the opacity of the atmosphere other than that involved in rain or cloud. The whole theory of the heat balance of the atmosphere is in a rudimentary state.

The correlations of terrestrial temperatures and rainfall with solar radiation are certainly small; a summary of the results obtained, with references to the literature, is given by C. E. P. Brooks in the *Meteorological Magazine* for June 1921||. A connexion between sunspots and cyclones in the South Indian Ocean has been traced by Meldrum, but cyclones in other oceans show little connexion with sunspots or with one another¶.

E. Huntington and S. S. Visher** have given an account of a theory of climatic changes, attributing most of the changes somewhat naïvely to

* *Geog. Journ.* 1923.

† *Zs. Osterr. Ges. Meteor.* **8**, 1873, 241–252.

‡ *Nature*, **43**, 1891, 583.

§ *Mems. Indian Met. Dept.* **24**, 1922—25, No. 4.

|| Cf. also *First Rept. of Commission on Solar and Terrestrial Relationships*, Paris, 1926.

¶ Mrs G. F. Newnham, *Geophys. Mems. of the Meteorological Office*, No. 19, 1922.

** *Climatic Changes, their Nature and Causes.* Yale Univ. Press, 1922.

variations in solar radiation. It certainly seems that some of the changes are most probably due to such variations, more on account of the failure of other hypotheses than on account of any conspicuous successes of this one. There is so far no satisfactory explanation of why the sun's radiation should have varied in the past; but as we do not know why it varies at present this offers no serious objection. Huntington and Visher offer an explanation based on the hypothesis that extra solar radiation is stimulated by variations in the radiation received by the sun from the stars; but this can be rejected, for if it were so, the components of a double star with a highly eccentric orbit should stimulate each other to such an extent as to make such stars strongly variable in a period equal to the period of revolution, which is not the case.

APPENDIX E

Empirical Periodicities

"Backward or forward, it's just as far;
Out or in, the way's as narrow." IBSEN, *Peer Gynt*.

E·1. Methods of harmonic analysis have been applied to many of the data of geophysics, especially in seismology, terrestrial magnetism, and meteorology. The principle of the method is as follows.

Suppose we have $2n+1$ observations of a variable x, distributed at equal intervals of another variable t. As a rule t is the time. We may measure t from the first observation. Let the interval be T. Then the value of t at the $(p+1)$th observation is pT, and at the last is $2nT$. Let the value of x when $t = pT$ be x_p. Put

$$\theta = \frac{2\pi t}{(2n+1)T} \qquad \text{.........(1).}$$

Consider the function

$$a_0 + a_1 \cos\theta + b_1 \sin\theta + \ldots + a_r \cos r\theta + b_r \sin r\theta + \ldots$$
$$+ a_n \cos n\theta + b_n \sin n\theta \qquad \text{.........(2),}$$

where the a's and b's are unknown constants, $2n+1$ in number. If we suppose this function equal to x, our $2n+1$ known values of x give $2n+1$ linear equations to determine these coefficients, which can therefore be found uniquely provided the equations are determinate. We notice that as t increases by T, θ increases by $2\pi/(2n+1) = \alpha$, say. The equation given by x_p is then

$$a_0 + a_1 \cos p\alpha + b_1 \sin p\alpha + \ldots + a_r \cos rp\alpha + b_r \sin rp\alpha + \ldots$$
$$+ a_n \cos np\alpha + b_n \sin np\alpha = x_p \ \text{...(3).}$$

There are equations of this type for all integral values of p from 0, 1 to $2n$. Multiplying them in turn by

$$\cos 0,\ \cos r\alpha,\ \cos 2r\alpha,\ \ldots\ \cos pr\alpha,\ \ldots\ \cos 2nr\alpha,$$

and adding, we see that the coefficient of a_s in the sum is

$$\sum_{p=0}^{2n} \cos rp\alpha \cos sp\alpha = \tfrac{1}{2} \sum_{p=0}^{2n} \{\cos (r-s)p\alpha + \cos (r+s)p\alpha\} \ \text{......(4).}$$

Several cases arise. If r and s are unequal, each sum is of the form

$$\tfrac{1}{2} \sum_{p=0}^{2n} \cos mp\alpha = \frac{\sin \dfrac{2n+1}{2} m\alpha \cos nm\alpha}{2 \sin \frac{1}{2} m\alpha} \qquad \text{.........(5),}$$

where m is an integer, and the first factor in the numerator is $\sin m\pi$, or zero. Thus in this case the coefficient is zero.

If r and s are equal but not zero, the second sum vanishes. The first is evidently $\frac{1}{2}(2n+1)$.

If r and s are both zero, each sum is $\frac{1}{2}(2n+1)$ and therefore the two together are equal to $2n+1$. Hence

$$(2n+1)\,a_0 = \sum_{p=0}^{2n} x_p \qquad \text{.........(6)},$$

$$\tfrac{1}{2}(2n+1)\,a_r = \sum_{p=0}^{2n} x_p \cos rpa \qquad \text{.........(7)}.$$

Similarly we can show that

$$\tfrac{1}{2}(2n+1)\,b_r = \sum_{p=0}^{2n} x_p \sin rpa \qquad \text{.........(8)}.$$

Thus all the coefficients are determinate, and we have found the coefficients $a_0 \ldots \ldots a_n$, $b_1 \ldots \ldots b_n$, so as to make the trigonometric function (2) equal to x for all the observed values of t. Thus any variable observed for $2n+1$ equidistant values of t can be represented exactly for these values of t by a series of $2n+1$ harmonic terms. This is the method known as harmonic or Fourier analysis.

E·2. We notice that the possibility of this representation is in no way dependent on the existence of any recognizable physical connexion between x and t. A completely random set of $2n+1$ observations, subjected to harmonic analysis, will give $2n+1$ harmonic terms. Thus if we determine the Fourier coefficients for any definite period, and find them different from zero, the result by itself gives no information that enables us to say that x is, even in part, related to t by any recognizable law.

If $\bar{x}$ be the mean value of x, and if the deviations of x from $\bar{x}$ were quite unrelated to $\cos r\theta$ and $\sin r\theta$, we should still find that the a's and b's did not vanish. This indeed is obvious; for if they did vanish, our theorem shows that x would be the same for all the observed values of t, which is not the case.

Now clearly $\bar{x} = a_0$. Let $x - a_0 = y$. Then

$$\begin{aligned}\sum_{p=0}^{2n} y_p^2 &= \sum_{p=0}^{2n} (a_1 \cos pa + b_1 \sin pa + \ldots + a_r \cos rpa + b_r \sin rpa + \ldots)^2 \\ &= (2n+1) \sum_{r=1}^{n} \tfrac{1}{2}(a_r^2 + b_r^2) \\ &\quad + \tfrac{1}{2} \sum_{p=0}^{2n} \sum_{r=1}^{n} (a_r^2 - b_r^2) \cos 2rpa \\ &\quad + \text{other trigonometric terms} \qquad \text{.........(9)}.\end{aligned}$$

All the trigonometric terms give zero on addition with regard to p, by (5). Also if σ denote the standard deviation of x, or of y, defined by

$$(2n+1)\,\sigma^2 = \sum_{p=0}^{2n} y_p^2 \qquad \text{.........(10)},$$

we have

$$\Sigma_r (a_r^2 + b_r^2) = 2\sigma^2 \qquad \text{.........(11)}.$$

But $a_r^2 + b_r^2$ is the square of the whole amplitude of the harmonic term (cosine and sine terms together) with period $(2n+1)\,T/r$. The number of terms other than the constant one is n. Thus the mean square amplitude of all the terms is

$$\sigma_1 = \sigma \sqrt{(2/n)} \qquad \text{.........(12)}.$$

This may be called the standard amplitude.

If a mere set of random observations, $2n+1$ in number, were analysed,

the mean square harmonic coefficient would have this value. The coefficients, however, will in general not be all equal, and several of them will exceed this mean. The principle of 'Schuster's criterion' is that a term can be considered as having physical meaning and not as being a result of the analysis of random fluctuations, if its amplitude is considerably greater than the standard amplitude*. Greater precision has been given to the criterion by Sir G. T. Walker†. The probability of a coefficient c_r given by

$$c_r^2 = a_r^2 + b_r^2 \qquad \text{.........(13)}$$

exceeding ξ is‡

$$e^{-\xi^2/\sigma_1^2} \qquad \text{.........(14).}$$

Two cases clearly arise. If there is *a priori* reason to expect a given period in the observations, and the observations when analysed show this period, they may be considered as confirming the earlier anticipations if the amplitude found is substantially greater than σ_1, so that it is too great to be likely to have arisen by chance. Thus we can derive the following table:

ξ/σ_1	$e^{-\xi^2/\sigma_1^2}$	ξ/σ_1	$e^{-\xi^2/\sigma_1^2}$
1	0·37	2·5	0·002
1·5	0·105	3·0	0·0001
2·0	0·018		

It appears that an amplitude greater than about $2\sigma_1$ is so unlikely to have arisen by chance as to afford strong confirmation of the suggested periodicity.

But when the observations have been analysed without reference to previous considerations, the amplitudes found are merely those that stand out most prominently in the resulting analysis; that is, out of the n harmonic terms calculable from $2n + 1$ observations the largest is deliberately selected. If we have 111 observations, so that $n = 55$, it will be the normal event for *one* of the amplitudes to reach $2{\cdot}0\sigma_1$ by chance, so that such a value for the largest amplitude gives no confirmation of its reality. Walker gives the following values for n and k, where n is the number of amplitudes determined and $k\sigma_1$ is the probable value of the largest:

n	k	n	k
2	1·33	40	2·42
4	1·63	100	2·68
6	1·79	200	2·86
8	1·89	400	3·03
10	1·98	1000	3·24
20	2·21		

A single period not predicted already is trustworthy if the amplitude is substantially greater than $k\sigma_1$.

E·21. The periods of all the harmonic terms determined in E·1 are submultiples of $(2n + 1)\,T$. Suppose now that x was really a harmonic function of $\lambda\theta$, where λ is between m and $m + 1$, m being an integer, and

* *Proc. Roy. Soc.* A, **61**, 1897, 455–465.

† *Q.J.R. Met. Soc.* **51**, 1925, 337–346.

‡ This can be proved from the results in Whittaker and Robinson's *Calculus of Observations*, 168–176, and is given, in a slightly different form, by Schuster and Walker.

let us examine the values of the harmonic coefficients that would be found. For this purpose it is sufficient to replace the finite summations by integrals. If we are given

$$x = a \cos \lambda\theta + b \sin \lambda\theta,$$

we shall have

$$\pi a_r = \int_0^{2\pi} x \cos r\theta d\theta,$$

$$\pi b_r = \int_0^{2\pi} x \sin r\theta d\theta,$$

whence

$$2\pi a_r = a\left[\frac{1}{\lambda - r} \sin 2\pi(\lambda - r) + \frac{1}{\lambda + r} \sin 2\pi(\lambda + r)\right]$$
$$+ b\left[\frac{1}{\lambda - r}\{1 - \cos 2\pi(\lambda - r)\} + \frac{1}{\lambda + r}\{1 - \cos 2\pi(\lambda + r)\}\right],$$

$$2\pi b_r = -a\left[\frac{1}{\lambda - r}\{1 - \cos 2\pi(\lambda - r)\} - \frac{1}{\lambda + r}\{1 - \cos 2\pi(\lambda + r)\}\right]$$
$$+ b\left[\frac{1}{\lambda - r} \sin 2\pi(\lambda - r) - \frac{1}{\lambda + r} \sin 2\pi(\lambda + r)\right].$$

The terms depending on $2\pi(\lambda + r)$ are evidently at most of order $\frac{1}{r}$, but those involving $2\pi(\lambda - r)$ may be of order unity. Hence if the harmonics considered are of high order, the coefficients of the largest terms may be replaced by

$$\pi a_r = \frac{\sin \pi(\lambda - r)}{\lambda - r}\{a \cos \pi(\lambda - r) + b \sin \pi(\lambda - r)\},$$

$$\pi b_r = \frac{\sin \pi(\lambda - r)}{\lambda - r}\{- a \sin \pi(\lambda - r) + b \cos \pi(\lambda - r)\},$$

and

$$\pi(a_r^2 + b_r^2)^{\frac{1}{2}} = \frac{\sin \pi(\lambda - r)}{\lambda - r}(a^2 + b^2)^{\frac{1}{2}}.$$

If then λ is equal to $r + \frac{1}{2}$, the two largest terms found by the harmonic analysis will be equal in amplitude. For other values, the values of the amplitudes found for the terms in $r\theta$ and $(r + 1)\theta$ will suffice to determine $(a^2 + b^2)^{\frac{1}{2}}$ and λ. Thus an extension of the method of harmonic analysis makes it possible to find approximately the amplitude and period of a harmonic variation, even though its period may not be a submultiple of $(2n + 1)T$. Thus the results of the harmonic analysis serve as a compendium whence the harmonic variations with periods intermediate between those used in the analysis can be inferred*.

E·3. Supposing the existence of a harmonic term to be established by analysis, an interpretation of this term is required. The character of this interpretation is very different in different cases. The simplest type is represented by a forced vibration where the external variation producing the motion is known beforehand to have a definite period. The phenomena of the tides provide an excellent example. The disturbing potentials due to the moon and sun have been expressed as the sum of a number of terms of different periods. Theoretically, each of these should produce variations in the height of the tide and the velocity of the current

* Turner, *Tables for facilitating the use of Harmonic Analysis*, 1913.

at any given place, with the period of the disturbing term that produces them; and the total height and velocity are the resultants of those due to the terms separately. Thus a harmonic analysis of the tidal observations at a given station enables each constituent separately to be found, and the results can afterwards be used to predict the tides at the station at any subsequent time. In this form the method presents no methodological difficulty: it is used only to determine the amplitudes and phases of terms that we already know must exist.

The next type in order of difficulty is the determination of the period of a free vibration. We may have previous reason to suspect the existence of a free vibration, and our theory may be so reliable that we can infer its period with considerable accuracy. If this is so, the determination of the amplitude and phase presents no more difficulty than in the case of a forced vibration. But in most geophysical phenomena the free periods are not accurately known beforehand. All we know is that a certain type of oscillation is possible, and perhaps we may know in addition the order of magnitude of its period. Thus to find the period we must carry out a harmonic analysis of the observations to find out what period agrees best with them. The analysis is complicated by the fact that free vibrations as a rule are affected by damping, the amount of the latter being usually unknown beforehand. This is innocuous in a forced ibration, for it does not affect the period, but only the amplitude and phase, which we find directly. But in free vibration it affects the period and makes the amplitude subject to a steady diminution. If a further disturbance regenerates the vibration, the new vibration will not necessarily be in the same phase as the old one, and therefore an analysis covering both may give quite incorrect estimates of the amplitude and even the period. Again, we may have no strong theoretical reason to believe that the free vibration we are seeking actually exists, for it may have been completely damped out, even if, indeed, it ever started. In such a case we cannot exclude the possibility that an empirical term with a period agreeing only roughly with our preliminary estimate may be due to some completely different cause. Again, it may happen that several periods, all of the correct order of magnitude, are disclosed by the analysis. If so, it will be difficult to say definitely which of them represents the predicted free vibration.

The next degree of difficulty occurs when we have no previous reason whatever to believe in the existence of a variation with a period comparable with that found. The 11-year period of sunspots is an example. Nobody knows why the number of spots on the sun should vary; we only know that it does. There is an equal period in the frequency of magnetic storms, and it is natural to infer from the agreement between the periods that the two phenomena are connected by a physical law. Now in such cases as these the actual variation is never simple harmonic; other periods and an irregular part are always present. If there is a physical connexion, we should expect it to hold, say, between the mean sunspot number for the month and the whole number of magnetic storms in the month, and not merely between the 11-yearly parts of the two variations. Thus it is natural to inquire why the method of harmonic analysis should in such cases be preferred to the method of correlation.

The answer appears to be as follows. The harmonic analysis enables us to find sets of harmonic terms that will represent any variation exactly. If then we analyse the two phenomena as in E·1, and consider the terms

of any two periods P and Q, the average values of the products of the terms of period P in the sunspot number into the terms of period Q in the number of magnetic storms are zero. Thus the combinations of terms of different periods contribute nothing to the correlation coefficient. On the other hand, if we take together the terms of period P in both, their products will as a rule not vanish when averaged over a long period, and they will therefore contribute to the correlation coefficient. Thus a harmonic analysis tells us all that the correlation coefficient can tell us; but it also tells us what part of the correlation is due to every periodic term separately. Further, it gives the ratios of the amplitudes and the lags of the separate terms. A knowledge of these quantities, and of the way they vary with the period, must be helpful in elucidating the nature of the connexion between the phenomena considered.

The harmonic analysis of a single series of observations, whose components are not to be compared with those of any other series, does not appear to be necessarily of any scientific value. The case of a predicted free vibration is hardly a case in point, since the results of the analysis are to be compared with theoretical considerations based on other data. In any investigation of a series of observations four cases may arise. First, it may happen that the existence of harmonic terms is predicted by theoretical considerations, and that such terms are found to give a good representation of the observations. In this case, typified by the tides, harmonic analysis achieves its greatest possible success. Second, it may be found that the observations are sufficient to establish a harmonic term or a few harmonic terms, but we may have no theoretical explanation of the existence of such terms. In this case the analysis supplies us with valuable data for future research, but not with knowledge relevant to any laws. Third, theory may predict the existence of a harmonic term, but on examination of the data it may be found that the harmonic term, though its existence may be well established, actually accounts for only a small fraction of the whole range of variation. This is specially liable to happen when long series of observations are analysed, for when n is large an amplitude may be both small compared with σ and large compared with σ_1. The lunar tide in the atmosphere, investigated by S. Chapman, affords a striking example. Such a result constitutes a verification of a theory, but it is evidently useless for prediction of individual observations. The fourth possibility is that we have no reason to expect the existence of a periodicity, and that on investigation we find that none of the amplitudes determined exceeds appreciably that expected from pure chance. In such a case we only find by analysis that nothing happens that we did not expect, which does not appear to be a result of any scientific interest. Many examples of such 'periodicities' have been published. The only means of distinguishing the second and fourth classes is Walker's criterion; but this criterion is seldom applied.

E·4. For our present purpose the most interesting periodicities are those in the frequencies of earthquakes. A great deal of work has been done on these periodicities, but so far only two appear to be well established. These are the solar diurnal and annual variations* found by

* C. Davison, *A Manual of Seismology*, Camb. Univ. Press, 1921, 82–198. References to papers are given.

C. Davison. It is found that, on the whole, earthquakes are more frequent in winter than in summer, and more frequent by night than by day.

It appears that these variations are of the nature of 'last straw phenomena.' Under the growing stresses in the earth's crust, fracture is bound to occur sooner or later. A small periodic variation in stress, however, may determine the time when the fracture actually takes place, provided the range of the variation is comparable with the amount of the steady increase during a period. If, for instance, the stress-differences in a region are greater, on the whole, in winter than in summer, they will be decreasing from winter to summer even when the steady increase is taken into account. Therefore there will be no earthquakes during this half-year; all will occur in the half of the year from summer to winter. Davison, largely following Omori, is inclined to attribute the annual period to the annual variation of atmospheric pressure, and the diurnal period to the diurnal variation of atmospheric pressure.

That these variations are not tidal in origin may be seen from the fact that the corresponding lunar terms are very much smaller, their amplitudes being little more than would be expected if the total variations were random. They have been the subject of investigation by R. D. Oldham*, and before him by C. G. Knott. The smallness of these lunar variations is, however, as interesting as their verification would have been. The extensional strain in the earth's crust since solidification has been seen to be of the order of 1 per cent., which has accumulated in a time of the order of 10^9 years. Thus the increase of extension in a month is of the order of 10^{-12}. The extension produced by the lunar bodily tide, however, must be of the order of 10^{-7}. It therefore becomes of interest to inquire why the lunar periodicity does not so dominate the occurrence of earthquakes as to make all the earthquakes in any region occur in only one half of the lunar day. The difficulty has been noticed by Oldham, who suggests that, while the gradual increase of stress produces set ultimately, the set is at first gradual and does not give a definite earthquake. When set has begun, however, it proceeds rapidly, and the small lunar effect is unable to influence appreciably the moment when the flow merges into fracture.

Other periodicities of much interest have been noticed by Prof. H. H. Turner. The aftershocks of earthquakes occur at intervals of multiples of 21 minutes†. Also the frequency of Chinese earthquakes, the heights of the Nile floods, the growth of Californian Sequoias, and the moon's longitude‡ all show a periodicity in a period of 250 to 300 years, those for the Californian trees being the best analysed because they provide the longest record. The records are such as to show such a period fairly definitely in all, and the phases agree as closely as the analysis can detect. No satisfactory explanation of the relations has been suggested.

* *Q. Journ. Geol. Soc.* **74**, 1918–19, 99–104; **77**, 1921, 1–3; **78**, 1922, lv–lxii.

† *M.N.R.A.S. Geoph. Suppl.* **1**, 1923, 31–50.

‡ H. H. Turner, *M.N.R.A.S.* **79**, 1919, 531–539; **80**, 1920, 617–619, 793–808. A. E. Douglass, *Climatic Cycles and Tree Growth*, Carnegie Institution Publ. No. 289, 1919, pp. 127.

INDEX

CAMBRIDGE: PRINTED BY W. LEWIS, M.A., AT THE UNIVERSITY PRESS